# 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

# 3D LEARNING IN ANATOMICAL AND SURGICAL EDUCATION IN RELATION TO VISUAL-SPATIAL ABILITIES

## *One size does not fit all*

*Katerina Bogomolova*

# 3D LEARNING IN ANATOMICAL AND SURGICAL EDUCATION IN RELATION TO VISUAL-SPATIAL ABILITIES

## *One size does not fit all*

Katerina Bogomolova

**3D LEARNING IN ANATOMICAL AND SURGICAL EDUCATION**
**IN RELATION TO VISUAL-SPATIAL ABILITIES**

ISBN: 978-94-6416-902-7

Cover design and layout: © evelienjagtman.com
Printing: Ridderprint

The printing of this thesis was financially supported by: Leiden University, Esser Foundation, the Dutch Association for Medical Education (NVMO), Incision Academy

# 3D LEARNING IN ANATOMICAL AND SURGICAL EDUCATION IN RELATION TO VISUAL-SPATIAL ABILITIES

## *One size does not fit all*

**Proefschrift**

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 3 februari 2022
klokke 10.00 uur

door

**Katerina Bogomolova**
geboren te Tashkent
in 1987

**Promotoren**
prof. dr. J.A. van der Hage
prof. dr. S.E.R. Hovius, Radboud University Medical Center, Nijmegen

**Copromotor**
dr. B.P. Hierck

**Promotiecommissie**
prof. dr. M.C. de Ruiter
prof. dr. M.E.J. Reinders, Erasmus University Medical Center Rotterdam
prof. dr. W.W. van den Broek, Erasmus University Medical Center Rotterdam
prof. dr. J.J.G. van Merriënboer, Maastricht University
prof. dr. H. van Goor, Radboud University Medical Center, Nijmegen

Посвящается моим родителям

Dedicated to my parents

# CONTENTS

# 1

# General Introduction

# GENERAL INTRODUCTION

## Anatomical knowledge

Anatomy has historically been a cornerstone in medical education. Appropriate knowledge of human anatomy is essential for medical specialties in general and surgery in particular. Mastering anatomical knowledge requires an accurate understanding of spatial relations of anatomical structures and the ability to translate this knowledge into practice. Unfortunately, changes in medical curricula have led to decreased teaching time for anatomy and limited exposure to traditional teaching methods such as dissections, prosections, and surface anatomy. As a result, students primarily learn from two-dimensional (2D) images in textbooks and anatomical atlases. Students experience difficulties in translating their 2D anatomical knowledge into practice. It is not surprising that the level of anatomical knowledge among medical students and junior doctors has been reported to be insufficient.[1-5] With reduced resident working hours, these challenges persist in surgical training as well. Surgical residents often feel unconfident performing surgeries characterized by spatial complexity.[6-8] This situation is imaginable because mastering spatially-complex procedures requires an accurate understanding of spatial relations of the relevant anatomical structures.

## Visual-spatial abilities

Translating anatomical knowledge into practice depends mainly on the level of visual-spatial abilities (VSA). In anatomical and medical contexts, VSA is defined as the ability to construct visual-spatial, e.g., three-dimensional (3D), mental representations of 2D images and mentally manipulate them.[9-10] Two components of VSA can be distinguished (Figure 1):

- Mental visualization and transformation: the ability to mentally construct and transform complex 3D objects
- Mental visualization and rotation: the ability to mentally rotate 3D objects and recognize them in other positions

VSA accurately predict the assessment of anatomical knowledge among students and surgical performance among residents, especially in the early phases of surgical training.[11,12] Not surprisingly, VSA is often recommended as a selection tool in surgical training.

The available assessment tools for VSA include psychometric tests that measure both components (Figure 1). The *Paper Folding Test (PFT)* is a standard test of mental visualization and transformation skills.[13] The test consists of ten parts in which subjects are asked to imagine the folding and unfolding of pieces of paper. After a hole has been punched in the folded piece of paper, the subject needs to indicate where the hole(s) will be after the pieces have been unfolded and the appearance of the unfolded pieces.

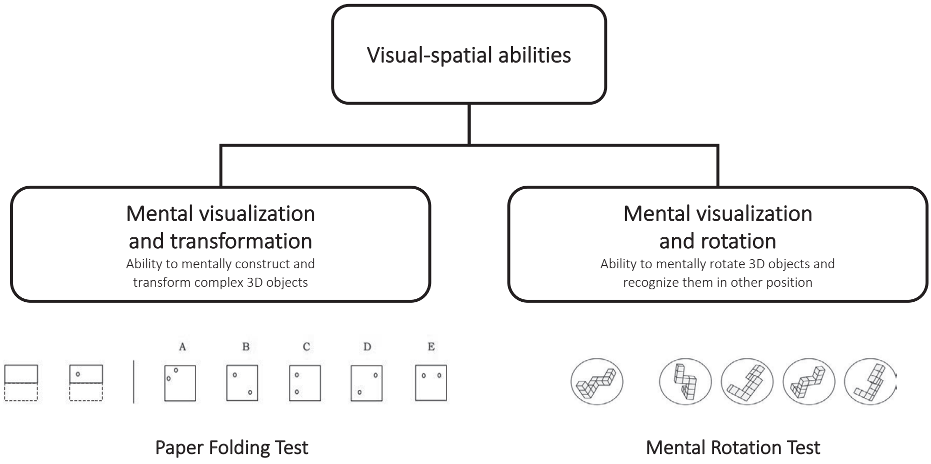


Figure 1. Components of visual-spatial abilities with examples of the most common used tests.

The *Mental Rotation Test (MRT)* is a validated test to assess mental rotation. This 24-item test has repeatedly showed significant positive associations with assessing anatomical knowledge and surgical skills.[14,15] Within each item, a 3D figure is presented as a 2D drawing with four possible rotated versions. Subjects must make a mental 3D representation of the figure and rotate it to identify the two correct options. One point is awarded for each correctly answered item, with a maximum score of 24 points.

### Three-dimensional visualization technology

Three-dimensional visualization technology (3DVT) can fill the gap between learning anatomy and applying the acquired knowledge in practice. The great advantage of 3DVT lies in its ability to visualize anatomical structures and explore spatial relations among structures from numerous viewpoints and angles. The '3D effect', or visual depth perception, in 3DVT is shaped by a mental combination of *monocular* and *binocular* depth cues.

*Monocular cues* are visual cues that require only one eye to perceive visual-spatial depth. They include coloring, relative size, shading, and motion parallax.[16] Here, we refer to 3DVT, which provides only monocular depth cues as *monoscopic 3DVT*. Examples of monoscopic 3DVT include 3D anatomical models, animations, and videos viewed on 2D displays such as computer screens, tablets, or cellular phones.

*Binocular cues* are visual cues that require both eyes to perceive visual depth. They include binocular disparity and convergence.[16] Binocular disparity is an essential cue for depth perception in real life. It refers to a slight difference between left and right retinal images of a 3D object derived from the eyes' horizontal distance. The disparity is detected by the brain and is translated into stereoscopic vision or stereopsis. Here, we refer to 3DVT, which provides binocular cues in addition to monocular depth cues as *stereoscopic 3DVT*. Stereoscopic vision in 3DVT is obtained by supportive devices that project two slightly different images to the left and right eye. These include autostereoscopic displays, anaglyphic or polarized 3D glasses, and head-mounted displays. Depending on the type of supportive devices, various stereoscopic 3D environments can be created, as follows.

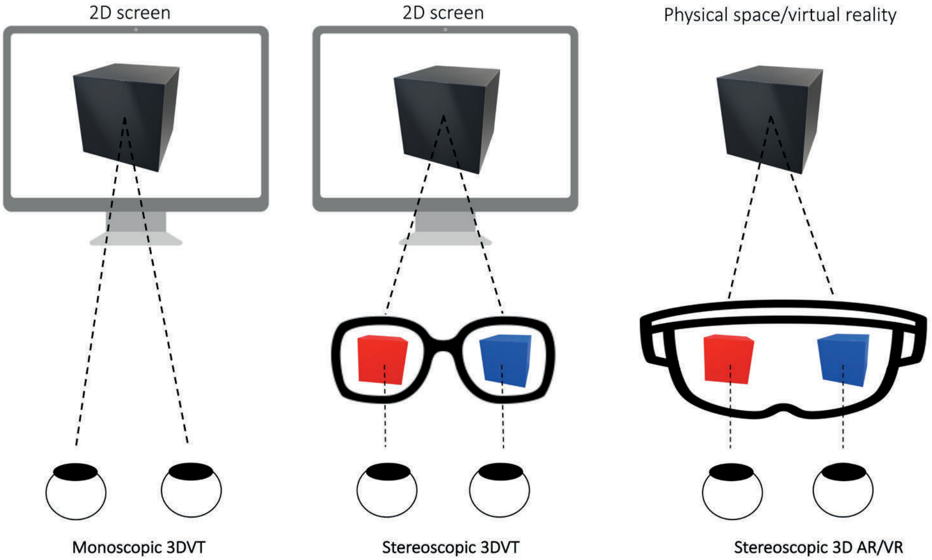


Figure 2. Monoscopic versus stereoscopic 3D environments. Stereoscopic vision is obtained by supportive devices that project two slightly different images to the left and right eye. 3DVT = three-dimensional visualization technology, AR = augmented reality, VR = virtual reality.

*Stereoscopic 3DVT* refers to the stereoscopic projection of 3D anatomical models, animations, or videos on 2D displays using anaglyphic, polarized, or shutter 3D glasses (Figure 2). Active interaction with the content is achieved by manual manipulation using a keyboard, computer mouse, or joystick. For 3D anatomical models, the user rotates the model, adjusts the size, and dissects anatomical structures layer-by-layer.

*Stereoscopic 3D augmented reality (AR)* refers to the stereoscopic projection of virtual 3D objects in natural environments using the HoloLens™, a pair of mixed reality smart glasses that run on the Windows operating system (Figure 2). The advantage of stereoscopic 3D AR is the ability to walk around the virtual object and explore it from all possible angles without losing the sense of one's physical environment. Active manipulation of the model includes rotation, size adjustment, relocation of the model in space, and dissection of structures layer-by-layer using hand gestures or voice commands. Multiple users wearing devices can share one model simultaneously. This type of technology is also referred to as interactive AR or mixed reality (MR).

*Virtual reality (VR)* refers to complete immersion into the virtual environment using VR head-sets such as Oculus Rift™ and HTC VIVE™ (Figure 2). VR allows the user to freely move in a virtual 3D space and explore 3D objects from all possible angles. Active manipulation of the model can include rotation, size adjustment, and dissection of structures layer-by-layer using motion-tracked handheld controllers.

### 3DVT in anatomical and surgical education

Although 3DVT may have distinct advantages for teaching and learning anatomy, its effectiveness remains a topic of continuing research and discussion. There are several reasons can be given to explain such a state of affairs.

First, most studies during the recent two decades focused on *whether* 3DVT works instead of *why* it works. In such cases, comparisons were made within levels of instructional design, meaning that the interventional method differed in numerous ways from the controlled method.[17] Therefore, the specific feature or mechanism of 3DVT that contributed to learning remained unknown. We must acknowledge that knowing whether a technology works is an essential step in the evaluation process of 3DVT. However, more importantly, research should focus on why a particular 3DVT works. The latter enables us to inform and advance medical education research and facilitate the implementation of 3DVT into daily educational practice.

Second, the learning effect of 3DVT appears not to be positive for all learners. Previous research has shown that monoscopic 3D desktop models have disadvantageous learning effects for students with low VSA.[17-21] Students with high VSA, however, do benefit from learning with monoscopic 3DVT. This situation brings two critical aspects into play. First, the disadvantageous learning effect of monoscopic 3DVT might be caused by the lack of accurate depth perception or stereoscopic vision. New technologies that provide good stereoscopic vision offer the potential to fill this gap. Second, VSA tends to cause a so-called aptitude-treatment interaction, or effect modification, in learning with various instructional methods.[18,20] This interaction occurs when the effect of an intervention appears to differ in groups of subjects with different characteristics (in our case, different levels of VSA). However, whether and how this VSA-induced interaction will occur in learning with stereoscopic 3DVT remains unknown.

In surgical education, 3DVT has the potential to be effectively used for learning and teaching surgical procedures. The need for additional teaching methods is increasing. According to the literature, surgical residents experience difficulties learning spatially-complex procedures and feel less confident performing such procedures despite proper preparation. Traditionally, surgical residents used surgical atlases and textbooks to prepare for surgeries; currently, they use online resources, including medical apps, books, and videos. Instructional videos are especially popular among residents.[22,23] This is not surprising because videos provide visual and auditory cues that facilitate mental visualization of procedural steps, including 3D aspects of anatomy. Nevertheless, the majority of research has focused on monoscopic visualization only. Whether a stereoscopic view of an instructional video would be more effective in preparing spatially-complex procedures remains unexplored.

Taking into consideration the above-mentioned aspects, the following questions arise:

1) How can we explain differences in learning with 3DVT between low- and high-VSA individuals?
2) Can we minimize these differences by providing stereoscopic vision in 3DVT?
3) Can VSA be taught to improve learning?

## Theoretical framework

To answer these questions, the fundamental concepts of this thesis are defined and explained within the theoretical framework of Cognitive Load Theory (CLT).[24] CLT provides instructional design principles and strategies based on the model of cognitive architecture. In this thesis, CLT served as the guide on which assumptions and predictions were built across the studies.

Figure 3. Human cognitive capacity: (a) cognitive overload, (b) preventing overload by decreasing extraneous load, (c) optimizing germane load (adapted from van Merrienboer & Sweller, 2010).[24]

The CLT model assumes that the human cognitive system has limited working memory capacity required to learn new information.[25] The task of working memory is to process novel information that is subsequently constructed into schemas in long-term memory. Three sources of working memory load can be distinguished: intrinsic cognitive load (caused by the nature of learning material content), extraneous cognitive load (caused by the way learning material is presented), and germane cognitive load (caused by actual learning process) (Figure 3).[25] When the sum of three sources exceeds the working memory capacity, cognitive overload occurs, consequently impairing the learning process. Based on this theory, individuals with low levels of VSA devote more cognitive resources to mental visualization and manipulation of 3D objects than individuals with high VSA levels. Therefore, when learning spatially-complex material (anatomy and surgical procedures), the intrinsic cognitive load in low-VSA individuals is elevated, leaving fewer available cognitive resources that they can spend on other learning tasks. To avoid cognitive overload, it is possible to decrease extraneous load by improving the instructional method, i.e., providing stereoscopic vision within 3DVT. The reduction in extraneous load will eventually provide more space for germane processing. With other words, teaching anatomy in ‘real 3D’ has the potential to improve actual learning by decreasing extraneous load.

## General aim and outline of the thesis

The overarching aim of this thesis was the employment of evidence-based insights to improve anatomical and surgical education by determining how various levels of VSA interact with learning using stereoscopic 3DVT.

In **chapter 2**, a meta-analysis was performed to estimate the learning effect of stereoscopic 3DVT compared to monoscopic 3DVT for learning anatomy. **Chapters 3 and 4** consider the effectiveness of stereoscopic 3D AR technology and its working mechanisms in learning the anatomy of the lower leg among (bio)medical students in relation to their VSA. The effect of stereoscopic vision was further explored in surgical procedure learning. In **chapter 5,** a randomized controlled trial was performed to evaluate the effect of stereoscopic 3D instructional video on the performance of a spatially-complex procedure among low- and high-VSA residents. In **chapter 6**, we evaluated how particular levels of VSA affect surgical performance depending on the type of intraoperative feedback. In **chapter 7**, we evaluated whether VSA can be trained and improved by the repeated anatomy practice. Finally, in **chapter 8,** the results of this thesis are put into a broader perspective and suggestions for future directions are made.

# REFERENCES

1. 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.
2. 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.
3. Spielmann PM, Oliver CW. The carpal bones: A basic test of medical students and junior doctors' knowledge of anatomy. Surgeon 2005;3:257–259.
4. Waterston SW, Stewart IJ. Survey of clinicians' attitudes to the anatomical teaching and knowledge of medical students. Clin Anat 2005;18:380–384.
5. Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. How much anatomy is enough? Anat Sci Educ 2008;1:184–188.
6. Klebanoff JS, Marfori CQ, Vargas MV, et al. Ob/Gyn resident self-perceived preparedness for minimally invasive surgery. *BMC Med Educ.* 2020;20(1):185.
7. Wanzel KR, Hamstra SJ, Anastakis DJ, et al. Effect of visual-spatial ability on learning of spatially-complex surgical skills. *Lancet.* 2002;359(9302):230-231.
8. Wanzel KR, Hamstra SJ, Caminiti MF, et al. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery.* 2003;134(5):750-757.
9. Gordon HW. The cognitive laterality battery: tests of specialized cognitive function. *Int J Neurosci.* 1986;29(3-4):223-244.
10. Kozhevnikov M, Hegarty M. A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem Cognit.* 2001;29(5):745-756.
11. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ 2017;10:235–241.
12. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and technical skills performance in health care: a systematic review. *Med Educ.* 2015;49(11):1065-1085.
13. Ekstrom RB, French J, Harman HH, Dermen D.. Kit of Factor-Referenced Cognitive Tests. Princeton, NJ: Educational Testing Service. 314 p. 1976.
14. Guillot A, Champely S, Batier C, Thiriet P, Collet C. Relationship between spatial abilities, mental rotation and functional anatomy learning. Adv Health Sci Educ Theory Pract 2007;12:491–507.
15. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ 2017;10:235–241.
16. Johnston EB, Cumming BG, Landy MS. Integration of stereopsis and motion shape cues. Vision Res. 1994;34(17):2259-2275.
17. Yammine K, Violato C. A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. Anat Sci Educ. 2015;8(6):525-538.
18. Luursema JM, Verwey WB, Kommers PAM, Annema JH. The role of stereopsis in virtual anatomical learning. Interact Comput. 2008;20(4-5):455-460.
19. Bogomolova K, van der Ham IJM, Dankbaar MEW, van den Broek WW, Hovius SER, van der Hage JA, Hierck BP. The effect of stereoscopic Augmented Reality visualization on learning anatomy and the modifying effect of visual-spatial abilities: a double-center randomized controlled trial. Anat Sci Educ. 2020; 13(5):558-567.
20. 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(1):34-45.

21. Huk T. Who benefits from learning with 3D models? The case of spatial ability. J Comput Assist Learn. 2006;22:392-404.
22. Green JL, Suresh V, Bittar P, et al. The utilization of video technology in surgical education: A systematic review. *J Surg Res.* 2019;235:171-180.
23. Ahmet A, Gamze K, Rustem M, Sezen KA. Is video-based education an effective method in surgical education? A systematic review. *J Surg Educ.* 2018;75(5):1150-1158.
24. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44(1):85-93.
25. Sweller J, van Merrienboer JJG, Paas F. Cognitive architecture and instructional design: 20 years later. *Educ Psychol Rev.* 2019;31(2):261-292.

# 2

# Stereoscopic three-dimensional visualization technology in anatomy learning:

## a meta-analysis

Katerina Bogomolova, Beerend P. Hierck, Agnes E. M. Looijen,
Johanne N.M. Pilon, Hein Putter, Bruce Wainman,
Steven E.R. Hovius, Jos A. van der Hage

*Medical Education 2021; 55:317-327*

# ABSTRACT

**Background**

The features that contribute to the apparent effectiveness of three-dimensional visualization technology (3DVT) in teaching anatomy are largely unknown. The aim of this study was to conduct a systematic review and meta-analysis of the role of stereopsis in learning anatomy with 3DVT.

**Methods**

The review was conducted and reported according to PRISMA Standards. Literature search of English articles was performed using Embase, Medline, CINAHL EBSCOhost, ERIC EBSCOhost, Cochrane CENTRAL, Web of Science and Google Scholar databases until November 2019. Study selection, data extraction and study appraisal were performed independently by two authors. Articles were assessed for methodological quality using the Medical Education Research Study Quality Instrument and the Cochrane Collaboration's tool for assessing the risk of bias. For quantitative analysis, studies were grouped based on relative between-intervention differences in instructional methods and type of control conditions.

**Results**

A total of 3934 citations were obtained of which 67 underwent a full-text review. Ultimately, 13 randomized controlled trials were included in the meta-analysis. When interactive, stereoscopic 3D models were compared to interactive, monoscopic 3D models within a single level of instructional design, e.g., isolating stereopsis as the only true manipulated element in the experimental design, an effect size (ES) of 0.53 (95% Confidence Interval (CI) 0.26 - 0.80; $p < .00001$) was found. In comparison with 2D images within multiple levels of instructional design, an effect size of 0.45 (95% CI 0.10 - 0.81; $p < .002$) was found. Stereopsis had no effect on learning when utilized with non-interactive 3D images (ES = - 0.87; 95% CI -2.09 - 0.35; $p = .16$).

**Conclusions**

Stereopsis is an important distinguishing element of 3DVT that has a significant positive effect on acquisition of anatomical knowledge when utilized within an interactive 3D environment. A distinction between stereoscopic and monoscopic 3DVT is essential to make in anatomical education and research.

# INTRODUCTION

Three-dimensional visualization technology (3DVT) is a promising tool in anatomy education. The first comprehensive summary and quantitative analysis of the effectiveness of 3DVT in teaching anatomy was performed by Yammine and colleagues in 2015.[1] In the meta-analysis, 3DVT interventions included combinations of technologies that allowed view of anatomy both in 3D (e.g., augmented and virtual reality) and two-dimensional (2D) environments (e.g., 3D models viewed on a 2D desktop computer).[1] It has been concluded that 3DVT has a positive effect on learning outcomes in terms of factual ($d = 0.30$) and spatial ($d = 0.50$) knowledge acquisition. However, to be able to implement this technology into educational practice, we need to know *why* this technology is effective. To do so, there are two important aspects that need to be addressed.

First, 3DVT appears to have disadvantages for students with low visual-spatial abilities (VSA) [1-5]. It has been hypothesized that digital multiple-view based, or 3D, images are being memorized as key-views based on familiar 2D images.[6,7] Consequently, when an unfamiliar 3D object is viewed from multiple angles, an increase in cognitive load occurs while generating a complete mental representation of a 3D object.[5] The proposed mechanism is in line with the ability-as-enhancer mechanism that is explained within the cognitive load theory.[8,9] According to this theory, individuals with higher VSA are able to devote more cognitive resources to building mental connections. Students with low VSA, on the other hand, get cognitively overloaded which eventually leads to underperformance.[5,7] However, according to additional research, when 3D models are presented stereoscopically, students with low VSA are able to reach the performance level of students with higher levels of VSA.[4] The observed opposite effect in the presence of stereopsis in 3DVT suggests its important and distinguishing role in learning.

Binocular stereopsis (also known as stereovision or stereo depth perception) is a result of binocular disparity between the right and left eye and this can be obtained in 3DVT by presenting slightly shifted 2D images to both eyes.[10] Stereovision can be produced with supportive devices such autostereoscopic displays (e.g., Alioscopy 3D Display [Alioscopy, Paris, FR]), anaglyphic or polarized glasses, or by a head-mounted display (e.g., HoloLens™ [Microsoft Corp., Redmond, WA, USA], or Oculus Rift™ [Oculus VR, Menlo Park, CA, USA] and HTC VIVE™ [High Tech Computer Corp., New Taipei City, Taiwan]). Hololens™ is used to create interactive augmented reality, also referred to as mixed reality (MR). Oculus Rift™ and HTC VIVE™ are predominantly used to create virtual reality (VR) environments. A binocular vision of the viewer, though, is required to perceive spatial visual depth that is obtained within this technology. Without stereopsis, the sense of visual depth in 3DVT is a result of a combination of monocular cues, such as shading, coloring, relative size and the motion parallax resulting from movement of the object.[11] In other words, there is no binocular disparity and thus only a monoscopic, or monocular, view of a 3D object results. Making a distinction between stereoscopic and monoscopic 3D visualizations is essential because it is a fundamentally different process. This critical nature of stereopsis is further supported by serial studies exploring the role of haptic feedback, transfer- appropriate processing and stereoscopic vision in the superiority of physical models above digital monoscopic 3D models.[12] Surprisingly, the large advantage of a physical model was predominantly due to stereoscopic vision, and not haptic feedback.

Second, within many studies, comparisons were made *between* levels of instructional designs (e.g., medium, configuration, instructional method, presentation), rather than *within* a single level.[13,14] Consequently, it remains unclear which element(s) or feature(s) of the interventions have contributed to the observed positive effect of 3DVT on learning. For instance, Codd and Choudhury have compared an interactive 3D model displayed in 2D of the upper limb with a combination of textbook and dissection.[15] Such comparison appears to be valid from a practical point of view since it resonates with the daily educational practice. However, it is unclear whether it was the configuration (monoscopic desktop view in contrast to 3D view in dissection including haptic feedback) or instructional method (self-regulated learning in contrast to a small group discussion during dissection) or both or some interaction of the two that contributed to the observed learning outcomes. Another common flaw in study designs is the inclusion of a control group with no training. Higher effect sizes will often be observed in favor of intervention when control group receives no 'treatment', as has been illustrated by several meta-analyses of Internet-based education.[16-19] As stated by Cook, such an effect appears logical, because if you teach students, they will eventually learn.[14] Therefore, inclusion of studies based on such comparisons in meta-analyses of educational effectiveness of 3DVT can lead to confounded outcomes and should be interpreted with caution.[20-24]

In the light of above considerations, the aim of this review was

(1) to provide a comprehensive summary of studies evaluating the educational effectiveness of stereoscopic 3DVT in anatomical education in relation to VSA
(2) to perform a meta-analysis to estimate the effect of stereopsis on anatomy learning by including studies with relatively few between-intervention differences in instructional methods (i.e., studies with comparisons made within one single level of instructional design).[13]

# METHODS

The review was conducted and reported according to PRISMA Standards of quality for reporting systematic reviews and best evidence medical education (BEME) collaboration methods.[25,26]

### Information sources and search strategy

Embase, Medline, CINAHL EBSCOhost, ERIC EBSCOhost, Cochrane CENTRAL, Web of Science and Google Scholar were searched for publications in English until November 2019. The search was augmented with manual searches in key journals and secondary screening through reference lists of existing reviews. The search strategy was conducted by the librarian and included following key terms: stereoscopic vision, three-dimensional, anatomical model and education.

### Eligibility criteria and study selection

Two independent reviewers (KB and AL or KB and AP) screened all titles and abstracts and excluded clearly irrelevant studies. The remaining articles underwent an independent, full- text screening by the same reviewers. Disagreements were solved through consensus. If consensus could not be reached, a third reviewer (BH) was consulted. The studies were selected according to the following hierarchical eligibility criteria:

1. Study was an original, full, peer-reviewed article written in English. *Conference papers, letters to editors, reviews, comments and study protocols were excluded.*
2. Study had an experimental comparative design including randomized controlled trials (RCT) and non-randomized comparative studies. *Studies with a single group with pretest and posttest, single group posttest only design and cross-sectional studies were excluded.*
3. Study subjects were university students in any academic field. *Studies that included high school students were excluded to avoid possible differences in levels of experience that can cause an expertise reversal effect.*[27]
4. Study intervention involved a teaching method with a stereoscopic 3D view of any anatomical region of the human or animal body. Stereoscopic 3D views could be obtained with the aid of any supportive device.
5. Control group involved any teaching method with a monoscopic view of the same anatomical region of the human or animal body. *Studies with control groups including non-digital teaching methods with a stereoscopic view such as cadaver or physical model, and control groups with no training were excluded.*
6. Study reported outcomes at level 4b of the Kirkpatrick's model, adopted by Steinert et al.[28], that included objectively assessed improvements of anatomical knowledge.

### Data extraction

Reviewers extracted the following data from each eligible study using a piloted extraction sheet: type of study design, target group and field, in/exclusion criteria (assessment of stereoscopic vision and VSA), number of participants, type(s) of educational intervention(s), anatomical region, type of assessment tool, outcome level, outcomes and their definitions, and cognitive level of questions. Cognitive levels of questions were categorized into low-order and high-order questions according to the Blooming Anatomy Tool.[29] This tool has been validated for use in educational research in anatomical sciences with improved consistency. Low-order questions were defined as reproduction of basic definitions and names of anatomical structures that only required information recall, and students were able to memorize the answers without understanding the process. In this case, questions that intended to assess spatial or functional understanding, but included identical images or texts from the study material, and thus stimulating only memorization and recall, were assigned as low-order questions. High-order questions were defined as transformation and application of acquired knowledge, including understanding of spatial organization, blood supply and innervation, functional anatomy, and applying information to a new situation or a new context. In this case, assessment images and text were different from the study material to ensure transformation and application of knowledge beyond memorization and recall. When information about the type of questions was insufficient, reviewer (KB) requested this information from authors via e-mail.

### Study appraisal

Methodological quality was assessed using the validated Medical Education Research Study Quality Instrument (MERSQI), that was developed for appraisal of the methodological quality of medical education research.[30] This assessment tool consists of ten items clustered in the following six domains: study design, sampling, type of data, validity of evidence for evaluation instrument scores, data analysis, and outcome. For each domain a minimum of 1 and maximum of 3 points could be awarded resulting in a total score ranging from 5 to 18.

Risk of bias was assessed using Cochrane Collaboration's tool for assessing risk of bias.[31] The tool includes seven domains: sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting and 'other issues'. For each domain 'low risk', 'high risk' or 'unclear risk' was assigned based on the criteria provided by the Cochrane Handbook. A 'high risk' of bias was assigned to the domain 'blinding of participants' if comparison was made between different types of media. An 'unclear risk' was assigned to seventh domain 'other issues' if stereoscopic vision of the participants was not assessed prior to the experiment.

## Data analysis

A descriptive analysis was used to summarize the included studies and to describe the effect of VSA on learning. A meta-analysis was performed to estimate the effect of stereopsis on learning outcomes. For the meta-analysis, studies were grouped based on relative between-intervention differences in instructional methods and type of control conditions (e.g., the ability of studies to isolate stereopsis as the only true manipulating element). This resulted in three types of comparisons:

*Interactive stereoscopic 3D models in comparison with interactive monoscopic 3D models.* The comparisons within each study were made within a single level of instructional design using the same medium and configuration. The only true element that differed between groups was the presence or absence of stereopsis. In the monoscopic view conditions, binocular disparity was avoided technically by presenting identical images to the left and right eyes or by covering the non-dominant eye of participants. Interaction included active manipulation of the model by the user (e.g., adjustment of the size, rotation of the model) and/or dynamic exploration (e.g., walking around the model) in case of interactive AR and VR environments.

*Interactive stereoscopic 3D models in comparison with 2D images.* The comparisons within each study were unavoidably made within one or more levels of instructional design using different types of medium and configuration. Therefore, stereopsis was not the only true manipulated element in the study design. Two dimensional images included non-interactive, monoscopic representations of anatomical structures on paper or a computer screen.

*Non-interactive stereoscopic 3D images in comparison with 2D images.* The comparisons within each study were made within a single level of instructional design using the same medium and configuration. The only true element that differed between groups was the presence or absence of stereopsis. Non-interactive stereoscopic 3D images included representations of anatomical 3D structures that could not be manipulated by the user and therefore perceived as static stereoscopic 3D images.

A sub-analysis was performed to evaluate the effect of stereoscopic 3DVT on the acquisition of factual and spatial knowledge domains separately.

The standardized mean differences (d), used as the effect size, were calculated based on given means and standard deviations. When insufficient information was provided, a given significance level was used to calculate the effect size. For studies with a pretest-posttest design we used posttest means. Heterogeneity between studies was quantified by $I^2$ statistics.[32] In case of large inconsistency, e.g., $I^2 > 50\%$, a random-effect model was used to pool the weighted effect sizes. Sensitivity analysis was performed by excluding studies with low methodological quality (MERSQI score <12) or with at least two or more assigned 'high risk' of bias. Publication bias was assessed using funnel plots and Egger's test.[33] Review Manager (Version 5.3, The Cochrane Collaboration, Oxford, England) was used for the analyses.

# RESULTS

### Study selection

The search strategy identified 3929 citations, and an additional 6 potentially relevant articles were identified from author files and review of reference lists (Figure 1). From these 69 potentially eligible articles were identified and a total of 16 studies were included in the qualitative synthesis. Three studies were excluded from the quantitative synthesis due to substantial between-intervention differences in instructional methods and type of control conditions, and a non-randomized study design.[34-36] Ultimately, 13 studies were included in the meta-analysis.

### Study characteristics

Among 16 studies included in the qualitative synthesis, study designs included randomized controlled trials with pre- and posttest (n = 7; 43.8%), posttest only (n = 7; 43.8%) and non-randomized comparative studies (n = 2; 12.4%) (*Supplementary material 1*). The included studies involved 1695 participants who were students in medicine (n = 422; 24.9%), nursing (n = 427; 25.2%), educational sciences (n = 420; 24.8%), medicine and biomedical sciences (n = 180; 10.6%), veterinary medicine (n = 84; 5.0%), behavioral sciences (n = 82; 4.8%), and combination of academic disciplines (n = 80; 4.7%). Stereovision of participants was assessed prior to the experiment and used as an inclusion criterion in four studies.[2,37-39] The most common anatomical regions studied were cerebrum and skull (n = 7; 41.3%) followed by abdomen (n = 3; 17.6%), head and neck (n = 2; 11.8%), pelvis (n = 2; 11.8%), cardiac and thorax anatomy (n = 2; 11.8%), and lower limb (n = 1; 5.9%). Several types of interventions were identified, including: stereoscopic 3D model with interactive user control (n = 11; 68.7%); stereoscopic 3D model with interactive instructor control (n = 1; 6.3%); and stereoscopic non-interactive 3D images (n = 4; 25.0%). Stereoscopic view was obtained with the aid of anaglyphic glasses 3D glasses (n = 4; 23.5%), 3D shutter glasses (n = 3; 17.6%), polarizing 3D glasses (n = 2; 11.8%), head-mounted displays such as Oculus Rift (n = 3; 17.6%), HTC Vive (n = 2; 11.8%) and HoloLens (n = 2; 11.8%), and autostereoscopic hologram i.e., the images was perceived in 3D without head-mounted device (n = 1; 5.9%).

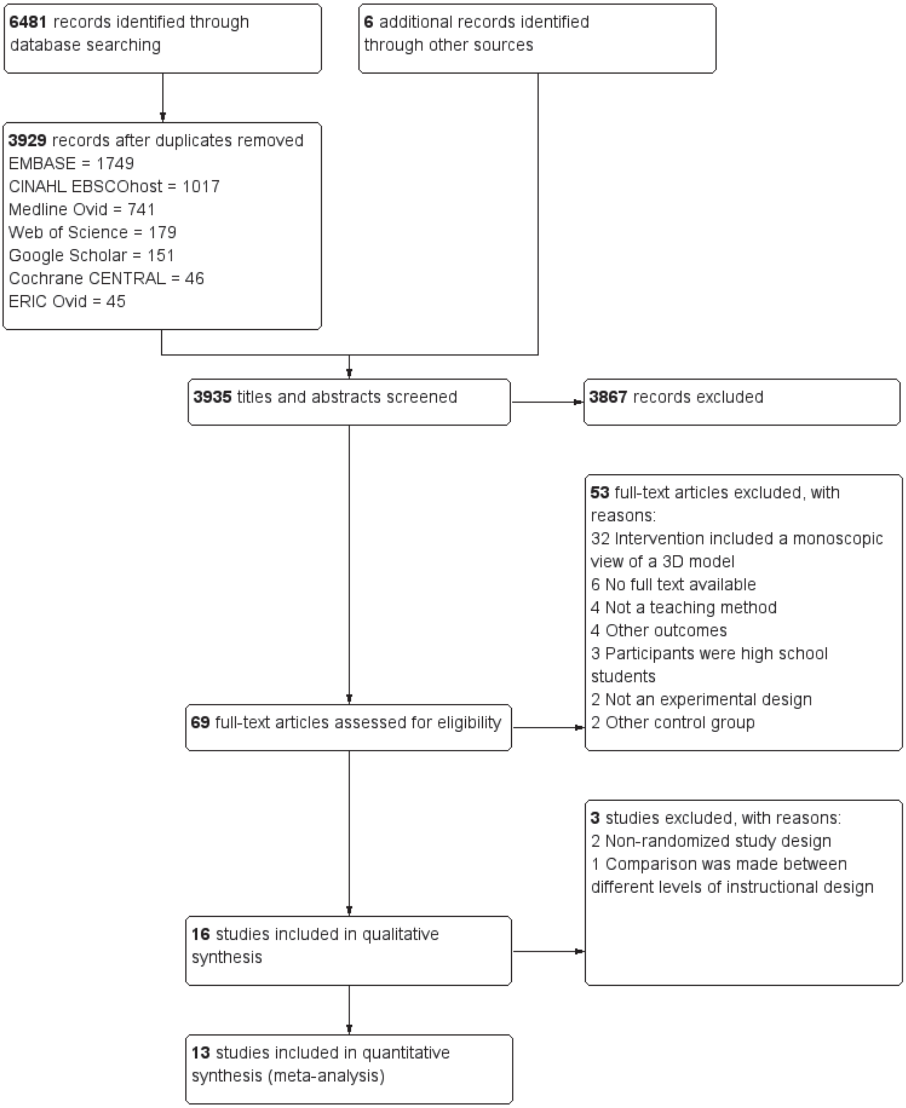


Figure 1. Flow diagram of study selection.

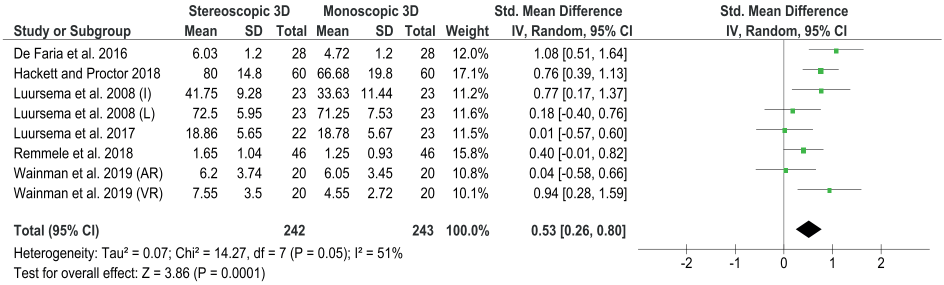
### Study appraisal

The mean MERSQI score for the 16 included studies was 13.0, ranging from 10.5 to 15. Reduction in scores and assigned 'high risk' of bias was primarily due to non-randomized study design of several studies. The MERSQI score for the studies included in the meta-analysis ranged from 12.5 and 15.0, with the highest score and no assigned 'high risk' of bias for the comparison 'interactive stereoscopic 3D model versus interactive monoscopic 3D model'.

## Meta-analysis

*Interactive stereoscopic 3D models in comparison with interactive monoscopic 3D models*

Six studies compared interactive stereoscopic 3D models with interactive monoscopic 3D models within a single level of instructional design.[2,37,39-42] One study evaluated two outcomes (i.e., identification task and localization task)[2] and one study evaluated two interventions (i.e., VR and interactive AR) separately.[42] A significant positive effect on overall anatomical knowledge was observed in favor of interactive stereoscopic 3D models (ES = 0.53, 95% CI 0.26 - 0.80; p < .00001; $I^2$ = 51%; n = 8) (Figure 2).



| | Stereoscopic 3D | | | Monoscopic 3D | | | | Std. Mean Difference |
|---|---|---|---|---|---|---|---|---|
| Study or Subgroup | Mean | SD | Total | Mean | SD | Total | Weight | IV, Random, 95% CI |
| De Faria et al. 2016 | 6.03 | 1.2 | 28 | 4.72 | 1.2 | 28 | 12.0% | 1.08 [0.51, 1.64] |
| Hackett and Proctor 2018 | 80 | 14.8 | 60 | 66.68 | 19.8 | 60 | 17.1% | 0.76 [0.39, 1.13] |
| Luursema et al. 2008 (I) | 41.75 | 9.28 | 23 | 33.63 | 11.44 | 23 | 11.2% | 0.77 [0.17, 1.37] |
| Luursema et al. 2008 (L) | 72.5 | 5.95 | 23 | 71.25 | 7.53 | 23 | 11.6% | 0.18 [-0.40, 0.76] |
| Luursema et al. 2017 | 18.86 | 5.65 | 22 | 18.78 | 5.67 | 23 | 11.5% | 0.01 [-0.57, 0.60] |
| Remmele et al. 2018 | 1.65 | 1.04 | 46 | 1.25 | 0.93 | 46 | 15.8% | 0.40 [-0.01, 0.82] |
| Wainman et al. 2019 (AR) | 6.2 | 3.74 | 20 | 6.05 | 3.45 | 20 | 10.8% | 0.04 [-0.58, 0.66] |
| Wainman et al. 2019 (VR) | 7.55 | 3.5 | 20 | 4.55 | 2.72 | 20 | 10.1% | 0.94 [0.28, 1.59] |
| **Total (95% CI)** | | | **242** | | | **243** | **100.0%** | **0.53 [0.26, 0.80]** |

Heterogeneity: Tau² = 0.07; Chi² = 14.27, df = 7 (P = 0.05); I² = 51%
Test for overall effect: Z = 3.86 (P = 0.0001)

Figure 2. Pooled effect size for studies comparing interactive stereoscopic 3D models with interactive monoscopic 3D models. 95% CI, 95% confidence interval; I, identification task; L, localization task; VR, virtual reality environment; AR, interactive augmented reality environment.

The funnel plot for the included studies showed no asymmetry, which suggests the absence of publication bias (Figure 3). Egger's test could not be performed due to a small number of studies (n < 10). In a sub-analysis, the pooled effect sizes for low and high order questions remain significant in favor of stereoscopic 3D models and were 0.71 (95% CI 0.31 – 1.11; p = .005; $I^2$ = 55%, n = 4) and 0.35 (95% CI 0.06 - 0.63; p = .02; $I^2$ = 14%; n = 4) respectively.

*Interactive stereoscopic 3D models in comparison with 2D images*

Seven studies compared interactive stereoscopic 3D models with 2D images.[3,4,38,40,41,43,44] A significant effect on anatomical knowledge was observed in favor of interactive stereoscopic 3D models (ES = 0.45, 95% CI 0.10 - 0.81; p < .002; $I^2$ = 70%; n = 8) (Figure 4). The funnel plot for the included studies showed some asymmetry suggesting the presence of a publication bias. Egger's test could not be performed due to a small number of studies (n < 10). A sub-analysis resulted in non-significant effects in favor of interactive stereoscopic 3D models in terms of low-order (ES = 0.32, 95% CI -0.18 - 0.81; p = .21; $I^2$ = 72%; n = 3) and high-order questions (ES = 0.73, 95% CI -0.03 – 1.49; p = .06; $I^2$ = 77%; n = 3).

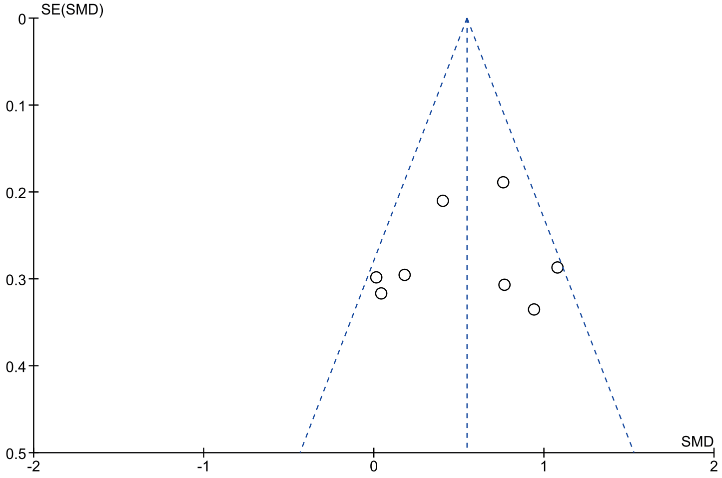
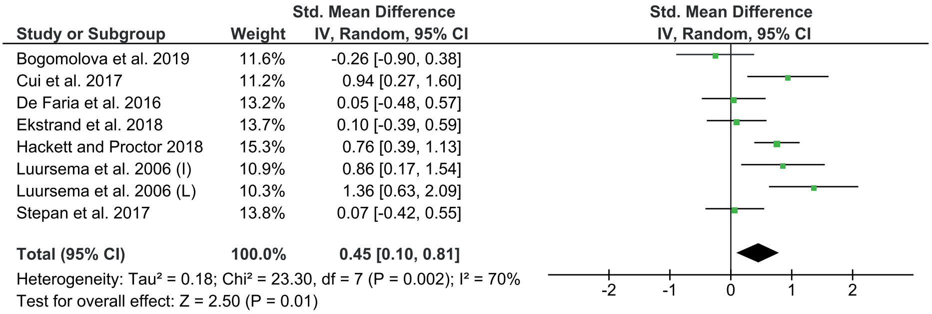


Figure 3. Funnel plot for studies included in the meta-analysis comparing interactive stereoscopic 3D models with interactive monoscopic 3D models.



| Study or Subgroup | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|
| Bogomolova et al. 2019 | 11.6% | -0.26 [-0.90, 0.38] |
| Cui et al. 2017 | 11.2% | 0.94 [0.27, 1.60] |
| De Faria et al. 2016 | 13.2% | 0.05 [-0.48, 0.57] |
| Ekstrand et al. 2018 | 13.7% | 0.10 [-0.39, 0.59] |
| Hackett and Proctor 2018 | 15.3% | 0.76 [0.39, 1.13] |
| Luursema et al. 2006 (I) | 10.9% | 0.86 [0.17, 1.54] |
| Luursema et al. 2006 (L) | 10.3% | 1.36 [0.63, 2.09] |
| Stepan et al. 2017 | 13.8% | 0.07 [-0.42, 0.55] |
| **Total (95% CI)** | **100.0%** | **0.45 [0.10, 0.81]** |

Heterogeneity: Tau² = 0.18; Chi² = 23.30, df = 7 (P = 0.002); I² = 70%
Test for overall effect: Z = 2.50 (P = 0.01)

Figure 4. Pooled effect size for studies comparing interactive stereoscopic 3D models with 2D images. 95% CI, 95% confidence interval; I, identification task; L, localization task.

*Non-interactive stereoscopic 3D images in comparison with 2D images*

Three studies compared non-interactive stereoscopic 3D images with 2D images within a single level of instructional design.[39,45,46] One study evaluated three anatomical regions (abdomen, pelvis, thorax) separately.[45] Meta-analysis showed a non-significant effect in favor of 2D images for overall anatomical knowledge (ES = - 0.87, 95% CI -2.09 - 0.35; p = .16; n = 5) (Figure 5). However, the $I^2$ of 97% indicated high heterogeneity of results between studies, and results could not be pooled. A sensitivity analysis did not change the heterogeneity across the studies. Because of the high heterogeneity, the funnel plot was not performed.[33]

| Study or Subgroup | Weight | Std. Mean Difference IV, Random, 95% CI |
|---|---|---|
| Al-Khalili et al. 2014 (A) | 20.2% | 0.00 [-0.52, 0.52] |
| Al-Khalili et al. 2014 (P) | 18.9% | -3.65 [-4.53, -2.78] |
| Al-Khalili et al. 2014 (T) | 19.8% | -1.89 [-2.52, -1.25] |
| Kockro et al. 2015 | 20.7% | -0.12 [-0.42, 0.18] |
| Remmele et al. 2018 | 20.3% | 1.09 [0.62, 1.56] |
| **Total (95% CI)** | **100.0%** | **-0.87 [-2.09, 0.35]** |

Heterogeneity: $Tau^2 = 1.84$; $Chi^2 = 116.80$, df = 4 ($P < 0.00001$); $I^2 = 97\%$
Test for overall effect: Z = 1.40 (P = 0.16)

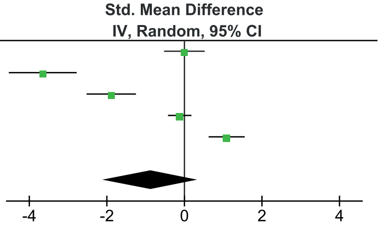


Figure 5. Pooled effect size for studies comparing non-interactive stereoscopic 3D images with 2D images. 95% CI, 95% confidence interval; A, abdomen; P, pelvis; T, thorax.

### The effect of VSA

VSA of participants were measured in six studies by a Mental Rotation Test (MRT). [2–4,37,38,42] Five studies used the redrawn version of MRT by Peters and colleagues[47] and one study[42] used the original version of Vandenberg and Kruse.[48] One study adjusted the outcomes for VSA by treating it as a confounder and reported that MRT scores significantly predicted performance (r = 0.548, P<0.0001).[42] Five studies evaluated the possible modifying effect of VSA by including VSA in a linear regression analysis as an interaction term[2,37,38] or by stratifying outcomes by VSA.[3,4] A significant interaction caused by VSA was reported in three studies.[3,4,38] This interaction meant that participants with lower VSA benefited significantly more from the stereoscopic view of anatomy than individuals with high VSA. Moreover, in the 3D stereoscopic group, differences between students with low and high levels of VSA were no longer significant.[3,4] One study reported similar interaction that did not reach a significant level (p = .09) in the linear regression analysis.[2] One study reported that the included interaction term was not significant and was therefore excluded from the regression analysis.[37] A meta-analysis of the modifying effect of VSA could not be performed due to insufficient data.

## DISCUSSION

The findings of the meta-analyses indicate that the presence of stereopsis, as a distinguishing feature of 3DVT, contributes to a better comprehension of anatomical knowledge. The beneficial effect of stereopsis (ES = 0.54) was observed when students learned anatomy using interactive 3D models that enabled active manipulation and/or dynamic exploration by the learner. The comparisons between stereoscopic and monoscopic interactive 3D models were made within a single level of instructional design, e.g., isolating stereopsis as the only true manipulated element in the experimental design. Similar effect was found when stereoscopic interactive 3D models were compared to 2D images (ES = 0.50). However, because the comparisons within each study were made between various levels of instructional design, it remains unclear to what extent stereopsis has contributed to this positive learning effect. In non-interactive representations, when stereoscopic 3D images were compared to 2D images within a single level of instructional design, stereopsis did not show any positive effect on learning. Heterogeneity among those studies was, however, large, and subgroup analyses did little to explain these inconsistencies.

The beneficial effect of stereopsis on learning anatomy, especially among students with low VSA, supports the hypothesis that the stereoscopic view contains spatial information not found in a monoscopic view that assists in generating an effective 3D mental representation of an object. This hypothesis suggests that with a monoscopic view, 3D objects are generated from key-view based 2D images rather than acquired naturally as a 3D object.[6,7] Generation of a 3D mental representation from memorized key-view 2D images requires certain amount of central processing which leads to an increase in cognitive load whenever an unfamiliar 3D object is viewed monoscopically compared to seeing the object stereoscopically. According to the ability-as-enhancer hypothesis, students with higher levels of VSA are able to allocate more cognitive resources to generate the required mental 3D representations and are able to benefit from monoscopic 3DVT.[9] Students with lower levels of VSA, however, lack this ability and experience difficulties to learn from monoscopic visualization. This explains why monoscopic 3DVT was found to have disadvantages for students with low VSA.[3-5,49]

The superiority of stereoscopic 3DVT over monoscopic 3DVT for students with low VSA implies another mechanism that is in line with the compensating hypothesis.[50] In stereoscopic 3DVT, the required 3D representation is already built and provided by stereoscopic vision. Consequently, generation of a mental 3D representation does not require additional mental steps leaving sufficient amount of cognitive resources to learn. In this way, stereoscopic view of a 3D object is able to compensate for low VSA students. A similar interaction in line of compensating hypothesis has been observed by Berney and colleagues comparing dynamic and static visualizations in learning anatomy.[51] Especially students with low VSA benefitted from learning with dynamic visualization. This was explained by the required mental representations of the movements that were already provided within dynamic visualizations, while representations whit static visualization still needed to be mentally genereated. Consequently, dynamic visualization was able to compensate for low VSA individuals.

The effect of stereopsis on learning is also being explored in the field of neuroeducational sciences. Anderson and colleagues have applied quantitative neural measures derived from electroencephalography to measure the effect of stereopsis in anatomy learning using a reinforcement-based learning paradigm.[52] When students learned anatomy using stereoscopic 3D models, greater object recognition was observed compared to those who learned from monoscopic 3D models. Another study in the field of neuroscience, has shown that binocular cues, in particular stereopsis, activate different neurons in the brain than monocular cues do.[53] Whether activation of different pathways in the brain would directly affect learning has not been demonstrated. However, both studies support that learning with monoscopic views of 3D objects is fundamentally different than learning with stereoscopic views.

The relationship between VSA and visualization type appears to be an aptitude-treatment interaction. An aptitude-treatment interaction occurs when a student attribute predicts different outcomes for different treatments.[13] In our case, students with low VSA benefited most from stereoscopic 3D models and showed a learning trajectory distince from the students with high VSA. It is important to mention that such interactions are only detectible when outcomes are stratified by the variable (VSA) or when the variable is included in the regression analysis as an interaction term. Including the variable only as a confounder will not reveal this interaction. Although a meta-analysis for various levels of VSA could not be performed due to an insufficient number of studies, these findings suggest, that VSA can potentially modify learning outcomes. Such an interaction caused by VSA has previously been reported in various contexts.[54,55] For statistical analysis and interpretation of the results, this means that when a new educational intervention has no better effect for high performing students, but it works well for the low performing students, the overall results will often overshadow these differences and the outcome will be 'no effect'. The 'no effect' will remain even after accounting for VSA by the study design (e.g., randomization) or statistical analysis (e.g., including it only as a covariate in the regression analysis). Therefore, for researchers it is essential to make the distinction between various levels of VSA by analyzing the outcomes for different groups of students separately or by including an interaction term in the linear regression analysis.[56,57]

Along with stereopsis, active user control appears to play an important role. The results of the current study showed a beneficial effect of stereopsis only in interventions involving active user control of 3D anatomical models. The importance of active user control in learning is described within the general framework of embodied cognition.[58] Active manipulation would lead to a more explicit connection between motor and visual process and, consequently, to a better learning. The effect of user control, or direct manipulation, in stereoscopic 3DVT has recently been described by Jang and colleagues.[55] Authors have found, that students, who were allowed to actively manipulate stereoscopic 3D model of the inner ear, performed significantly better on the posttest than students who passively watched the interaction in the same stereoscopic 3D environment. The findings were in line with supporting theory from the field of embodied cognition suggesting that direct manipulation of structures in a virtual environment can facilitate embodied representations of 3D structures. Another distinguishing feature that is more often available in the 3DVT technology is the ability to perform dynamic exploration. Being able to walk around the model with its own reference point can create an additive sense of depth. Further research is needed to evaluate to what extent the combination of these features contribute to learning outcomes.

A small but growing portion of population appear to have suboptimal stereoacuity.[59] An even smaller portion lacks stereovision entirely and , therefore, cannot perceive the obtained spatial visual depth by stereoscopic 3DVT. The precise prevalence remains unknown, since the numbers vary greatly between studies[60,61] and between methods of measuring stereoacuity.[59] In the current review, eleven (7%) out of 145 participants, that were screened for stereoacuity in four studies, were reported to have no stereoscopic vision and were excluded from the studies.[2,37-39] Whether this percentage can be extrapolated to studies that did not screen participants for their stereovision, is doubtful. Three of the four studies included study samples from the same geographic area in the Netherlands. Also different stereo tests were used including the TNO Random-dot test[2,38], Random Dot 3 LEA SYMBOLS Stereoacuity Test[37] and the Titmus stereotest[39] that can produce different results. Whether the degree of their stereoacuity was sufficient to perceive the disparity in the images presented in the 3DVT remains unknown.

### Limitations

The current meta-analysis used an approach focusing on the comparisons that were made within a single level of instructional design. This enabled a measure of the effect of stereopsis to be the only true manipulated element in the experimental design. Inherently, this approach has several limitations. First, no comparisons were made between various types of 3D technology, but instead, the effect of stereopsis was estimated within different types of 3DVT combined. It is possible for a particular type of technology to obtain a slightly different quality of stereopsis, and, therefore, affect learning experience. Also different side effects can occur such as blurred vision, headache and dizziness depending on the technology employed.[36] Second, no distinction was made between different anatomical regions that might require less or more VSA of learners. Third, no distinction was made between different ways of obtaining monoscopic views in control conditions. One study created a monosocpic view by covering the non-dominant eye of participants. This could have resulted in a different viewing condition compared to those presenting identical images to both eyes where some binocular cues could have remained in addition to monocular cues. Due to a relatively small number of studies, the described above distinctions could not be made within the meta-analyses. Additionally, publication bias could not be fully assessed for all comparison groups. For the main comparison group a funnel plot suggested no publication bias (Figure 3). However, for other groups the funnel plot was asymmetrical or could not be performed due to a high heterogeneity. This suggests that selective reporting may have cause an overestimation of effect sizes in small studies. Last, the subgroup analyses for low- and high-order anatomy questions should be interpreted with caution because of the number of comparisons made and the heterogeneity in anatomy knowledge tests. The content validation of anatomy knowledge tests was performed in 10 of the 16 studies, while the internal consistency of the tests was assessed in none of the studies.

## CONCLUSIONS

This was the first systematic review and meta-analysis evaluating the educational effect of stereopsis in 3DVT for teaching anatomy. Technically, stereoscopic view of a digital 3D object is different from a monoscopic view due to the projection of a slightly different image to the left and right eye resulting in a sense of a perceived depth. Therefore, it is essential to make a distinction between stereoscopic and monoscopic 3DVT in anatomical education and research. When designing new research, VSA and stereovision of learners should always be taken into account. From educational point of view, as supported by the results of this study, stereoscopic 3DVT contributes to a better comprehension of anatomy and is preferred over the monoscopic 3DVT, especially when utilized in an interactive 3D environment.

# REFERENCES

1. Yammine K, Violato C. A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. Anat Sci Educ 2015;8(6):525-538.
2. Luursema JM, Verwey WB, Kommers PAM, Annema JH. The role of stereopsis in virtual anatomical learning. Interact Comput 2008;20(4-5):455-460.
3. Bogomolova K, van der Ham IJM, Dankbaar MEW, van den Broek WW, Hovius SER, van der Hage JA, Hierck BP. The effect of stereoscopic Augmented Reality visualization on learning anatomy and the modifying effect of visual-spatial abilities: a double-center randomized controlled trial. Anat Sci Educ 2020; 13(5):558-567.
4. 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(1):34-45.
5. Huk T. Who benefits from learning with 3D models? The case of spatial ability. J Comput Assist Learn 2006;22:392-404.
6. Bulthoff HH, Edelman SY, Tarr MJ. How are three-dimensional objects represented in the brain? Cereb Cortex 1995;5(3):247-260.
7. Garg AX, Norman G, Sperotable L. How medical students learn spatial anatomy. Lancet 2001;357(9253):363-364.
8. Mayer RE. Multimedia Learning. Cambridge: Cambridge University Press; 2001.
9. Hegarty M, Sims VK. Individual differences in use of diagrams as external memory in mechanical reasoning. Memory & Cognition 1994;22(4):411-430.
10. Cutting JE, Vishton PM. Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In: Epstein W, Rogers, S, ed. Perception of Space and Motion. San Diego: Academic Press; 1995:69–117.
11. Johnston EB, Cumming BG, Landy MS. Integration of stereopsis and motion shape cues. Vision Res 1994;34(17):2259-2275.
12. 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(11):1138-1146.
13. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med 2005;80(6):541-548.
14. Cook DA. If you teach them, they will learn: why medical education needs comparative effectiveness research. Adv Health Sci Educ Theory Pract 2012;17(3):305-310.
15. Codd AM, Choudhury B. Virtual reality anatomy: is it comparable with traditional methods in the teaching of human forearm musculoskeletal anatomy? Anat Sci Educ.2011;4(3):119-125.
16. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med.2010;85(10):1589-1602.
17. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. JAMA 2008;300(10):1181-1196.
18. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA 2011;306(9):978-988.
19. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med 2011;86(6):706-711.

20. Abid B, Hentati N, Chevallier JM, Ghorbel A, Delmas V, Douard R. Traditional versus three-dimensional teaching of peritoneal embryogenesis: a comparative prospective study. Surg Radiol Anat 2010:1-6.
21. Hariri S, Rawn C, Srivastava S, Youngblood P, Ladd A. Evaluation of a surgical simulator for learning clinical anatomy. Med Educ 2004;38(8):896-902.
22. Petersson H, Sinkvist D, Wang C, Smedby O. Web-based interactive 3D visualization as a tool for improved anatomy learning. Anat Sci Educ 2009;2(2):61-68.
23. Seixas-Mikelus SA, Adal A, Kesavadas T, et al. Can image-based virtual reality help teach anatomy? J Endourol 2010;24(4):629-634.
24. Beermann J, Tetzlaff R, Bruckner T, et al. Three-dimensional visualisation improves understanding of surgical liver anatomy. Med Educ 2010;44(9):936-940.
25. Hammick M, Dornan T, Steinert Y. Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. Med Teach 2010;32(1):3-15.
26. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009;62(10):e1-34.
27. Kalyuga S, Rikers R, Paas F. Educational Implications of Expertise Reversal Effects in Learning and Performance of Complex Cognitive and Sensorimotor Skills. Educ Psychol Rev 2012;24(2):313-337.
28. Steinert Y, Mann K, Anderson B, et al. A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. Med Teach 2016;38(8):769-786.
29. Thompson AR, O'Loughlin VD. The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. Anat Sci Educ 2015;8(6):493-501.
30. Cook DA, Reed DA. Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. Acad Med 2015;90(8):1067-1076.
31. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
32. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-560.
33. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002.
34. Goodarzi A, Monti S, Lee D, Girgis F. Effect of Stereoscopic Anaglyphic 3-Dimensional Video Didactics on Learning Neuroanatomy. World Neurosurg 2017;107:35-39.
35. Hilbelink AJ. A measure of the effectiveness of incorporating 3D human anatomy into an online undergraduate laboratory. British Journal of Educational Technology 2009;40(4):664-672.
36. Moro C, Stromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017;10(6):549-559.
37. Luursema JM, Vorstenbosch M, Kooloos J. Stereopsis, Visuospatial Ability, and Virtual Reality in Anatomy Learning. Anat Res Int 2017;Article ID:1493135.
38. Luursema J, Verwey WB, Kommers PAM, Geelkerken RH, Vos HJ. Optimizing conditions for computer-assisted anatomical learning. Interact Comput 2006(18):1123–1138.

39. Remmele M, Schmidt E, Lingenfelder M, Martens A. The impact of stereoscopic imagery and motion on anatomical structure recognition and visual attention performance. Anat Sci Educ 2018;11(1):15-24.
40. de Faria JW, Teixeira MJ, de Moura Sousa Junior L, Otoch JP, Figueiredo EG. Virtual and stereoscopic anatomy: when virtual reality meets medical education. J Neurosurg 2016;125(5):1105-1111.
41. Hackett M, Proctor M. The effect of autostereoscopic holograms on anatomical knowledge: a randomised trial. Med Educ 2018;52(11):1147-1155.
42. Wainman B, Pukas G, Wolak L, Mohanraj S, Norman GR. The Critical Role of Stereopsis in Virtual and Mixed Reality Learning Environments. Anat Sci Educ 2020;13:401-412.
43. Ekstrand C, Jamal A, Nguyen R, Kudryk A, Mann J, Mendez I. Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: a randomized controlled study. CMAJ Open 2018;6(1):E103-E109.
44. Stepan K, Zeiger J, Hanchuk S, et al. Immersive virtual reality as a teaching tool for neuroanatomy. Int Forum Allergy Rhinol 2017;7(10):1006-1013.
45. Al-Khalili SM, Coppoc GL. 2D and 3D stereoscopic videos used as pre-anatomy lab tools improve students' examination performance in a veterinary gross anatomy course. J Vet Med Educ 2014;41(1):68-76.
46. Kockro RA, Amaxopoulou C, Killeen T, et al. Stereoscopic neuroanatomy lectures using a three-dimensional virtual reality environment. Ann Anat 2015;201:91-98.
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 Cogn 1995;28(1):39-58.
48. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills 1978;47(2):599-604.
49. Grag AX, Norman G, Spero L, Maheshwari P. Do virtual computer models hinder anatomy learning? Acad Med 1999;74(10):S87-S89.
50. Mayer RE, Sims VK. For Whom Is a Picture Worth 1000 Words - Extensions of a Dual-Coding Theory of Multimedia Learning. J Educ Psychol 1994;86(3):389-401.
51. Berney S, Betrancourt M, Molinari G, Hoyek N. How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models. Anat Sci Educ 2015;8(5):452-462.
52. Anderson SJ, Jamniczky HA, Krigolson OE, Coderre SP, Hecker KG. Quantifying two-dimensional and three-dimensional stereoscopic learning in anatomy using electroencephalography. NPJ Sci Learn 2019;4:10.
53. Verhoef BE, Vogels R, Janssen P. Binocular depth processing in the ventral visual pathway. Philos Trans R Soc Lond B Biol Sci 2016;371(1697).
54. Bogomolova K, Hierck BP, van der Hage JA, Hovius SER. Anatomy Dissection Course Improves the Initially Lower Levels of Visual-Spatial Abilities of Medical Undergraduates. Anat Sci Educ 2020;13(3):333-342.
55. Jang S, Vitale JM, Jyung RW, Black JB. Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. Computers & Education 2017;106:150-165.
56. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia:: Lippincott Williams & Wilkins; 2008.

57. Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandenbroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. Clin Epidemiol 2017;9:331-338.
58. Clark A. An embodied cognitive science? Trends Cogn Sci 1999;3(9):345-351.
59. Deepa BMS, Valarmathi A, Benita S. Assessment of stereo acuity levels using random dot stereo acuity chart in college students. J Family Med Prim Care 2019;8(12):3850-3853.
60. Birch E, Williams C, Drover J, et al. Randot Preschool Stereoacuity Test: normative data and validity. J AAPOS 2008;12(1):23-26.
61. Zaroff CM, Knutelska M, Frumkes TE. Variation in stereoacuity: normative description, fixation disparity, and the roles of aging and gender. Invest Ophthalmol Vis Sci. 2003;44(2):891-900.

# 3

# The effect of stereoscopic augmented reality visualization on learning anatomy and the modifying effect of visual-spatial abilities:

## a double-center randomized controlled trial

Katerina Bogomolova, Ineke J.M. van der Ham, Mary E.W. Dankbaar,
Walter W. van den Broek, Steven E.R. Hovius, Jos A. van der Hage, Beerend P. Hierck

*Anatomical Sciences Education 2020; 13:558-567*

# ABSTRACT

**Background**

Monoscopically projected three-dimensional (3D) visualization technology may have significant disadvantages for students with lower visual-spatial abilities (VSA) despite its overall effectiveness in teaching anatomy. Previous research suggests that stereopsis may facilitate a better comprehension of anatomical knowledge. This study evaluated the educational effectiveness of stereoscopic Augmented Reality (AR) visualization and the modifying effect of VSA on learning.

**Methods**

In a double-center randomized controlled trial, first and second-year (bio)medical undergraduates studied lower limb anatomy with a stereoscopic 3D AR model (n = 20), a monoscopic 3D desktop model (n = 20) or two-dimensional (2D) anatomical atlas (n = 18). VSA were tested with Mental Rotation Test (MRT), Paper Folding Test and Mechanical Reasoning Test. Anatomical knowledge was assessed by the validated 30-item paper post-test.

**Results**

The overall post-test scores in the stereoscopic 3D AR group (47.8 %) were similar to those in the monoscopic 3D desktop group (38.5 %; p = .081) and the 2D anatomical atlas group (50.9 %; p = 1.00). When stratified by VSA test scores, students with lower MRT scores achieved higher post-test scores in the stereoscopic 3D AR group (49.2 %) as compared to the monoscopic 3D desktop group (33.4 %; p = .015) and similar to the scores in the 2D group (46.4 %; p = .99). Participants with higher MRT scores performed equally well in all conditions.

**Conclusions**

It is instrumental to consider an aptitude-treatment interaction caused by VSA when designing research into 3D learning. Further research is needed to identify contributing features and the most effective way of introducing this technology into current educational programs.

# INTRODUCTION

Anatomical knowledge among undergraduate medical students and recently graduated doctors has repeatedly been reported to be insufficient.[1–5] One of the main reasons is the decrease in anatomy teaching time in undergraduate education, related to the increasing costs and limited availability of cadavers, and the time pressure on the curriculum have led to a decreased exposure to traditional cadaveric dissections.[4,6–9] Although, the educational value is being debated[7], cadaveric dissections provide a complete visual and tactile learning experience of anatomy which is three-dimensional (3D) by nature. Features such as stereopsis (visual sense of depth), dynamic exploration (the possibility to view the object of study from different angles), and haptic feedback (sense of touch) are crucial for the engagement in 3D anatomy.[8,9]

In search of additional educational resources, computer assisted resources have been widely explored in anatomical education. A considerable number of studies have evaluated the effectiveness of digital 3D anatomical models which can be explored on a two-dimensional (2D) screen of a regular computer, smartphone, or tablet. In an extended meta-analysis of these studies, Yammine and Violato[10] concluded that three-dimensional visualization technology (3DVT) is effective in improving factual (effect size of 0.50) and spatial (effect size of 0.30) anatomical knowledge. However, despite of the overall positive effect on learning, 3DVT appears to have significant disadvantages for students with low visual-spatial abilities (VSA).[11–15] The disadvantages are well known in the research field of 3D learning and were first described by Garg and colleagues. [11–13] In these studies, VSA significantly affected the learning process of spatial anatomy showing a great disadvantage for low performing students. Viewing an unfamiliar 3D object from multiple angles would be challenging for these students due to evidence that 3D objects are remembered as key view based 2D images.[13,14,16,17]

However, when traditional digital 3D models are viewed stereoscopically by projecting a slightly shifted image to the left and right eye, the disadvantages for low VSA students seem to disappear. Cui and colleagues have evaluated the effectiveness of a stereoscopic 3D view of the head and neck vascular anatomy in comparison to 2D representations of the same anatomical model.[18] They reported a better performance of undergraduate medical students after learning anatomy with a stereoscopic 3D model. Most importantly, students with low VSA have improved their knowledge test scores to a level comparable to that demonstrated by the high VSA students. The role of stereopsis has also been evaluated by Luursema and colleagues[19-21] within various 3D environments, such as virtual reality and stereoscopic projection on a computer with the use of 3D shutter glasses. Although the stereoscopic view of an anatomical model has had a positive effect only on one of the two post-tasks, the interaction between VSA and the stereoscopic condition remained significant.[20] Overall, stereoscopic 3DVT appears to have a positive effect on learning as recently demonstrated by Hackett and Proctor.[22] Their intervention concerned an autostereoscopic holographic visualization of a cardiac 3D model which has been compared to a monoscopic desktop view and 2D printed images of the model. Students in the intervention group scored significantly higher on the anatomical knowledge test and have reported a significantly lower cognitive load in comparison to both control groups. However, a possible interaction between intervention and VSA has not been evaluated. The positive role of stereopsis has also been shown when a physical model of the pelvis was compared to a monoscopic 3D model by Wainman and colleagues.[23] Authors have concluded that stereopsis, and not haptic feedback, primarily contributed to the improved knowledge scores when learning with a physical model.

In regard to these findings, two aspects come into play. Firstly, beneficial effects of stereopsis support the evidence that 3D mental representations depend on the nature of the input by activating different regions of the brain, and might contain spatial information instead of key-view based 2D images alone.[19,24-26] Stereopsis might therefore facilitate a better comprehension of anatomy especially among students with lower levels of VSA. Secondly, the reported differences in learning effect between students with lower and higher levels of VSA in various interventions possibly reflect an aptitude–treatment interaction. An aptitude-treatment interaction occurs when a student's attribute, e.g., visual-spatial abilities, predicts different outcomes for different treatments.[27] Such interaction is only detectible when the outcomes are stratified by the variable or when the variable is included in the regression analysis as an interaction term (variable x intervention), as demonstrated by Luursema and colleagues and Cui and colleagues.[18,20]

### Augmented reality in anatomical education

Augmented reality (AR) is a new generation of 3DVT technology that is eagerly being explored in the field of anatomical education and research in recent years.[27,28] It gained popularity due to its ability to combine 3D computer-generated virtual objects with physical environment. This enables learners to interact with each other and with the digital environment using mobile devices, such as smart phones and tablets, or, more recently, head-mounted displays (HMD) such as AR and virtual reality (VR) devices. Whether the anatomy can be perceived in a real three-dimensional plane, depends on the type of device. Visualization of 3D content from flat screens is usually obtained monoscopically with various interactive features added to the digital overlay provided by these devices.[29-31] HMD can provide an interactive and stereoscopic way of 3D visualization (*Supplementary material 2*). With AR technology, such as with the Hololens®, the most distinguishing feature is the ability to perceive an anatomical model in a real three-dimensional plane without losing sense of the user's own environment. Dynamic exploration, an object centered view, enables users to walk around the stereoscopic model and explore it from all possible angles. The use of this technology has been reported in the surgical field of preoperative planning and tumor localization.[32] The educational effectiveness of this technology for teaching anatomy has not been evaluated yet. For the purpose of this study an AR application *DynamicAnatomy* was developed at the department of Anatomy and Embryology at Leiden University Medical Center and the Centre for Innovation of Leiden University. This application provides a dynamic stereoscopic 3D view on the lower limb including the musculoskeletal anatomy. Further specification of the application is provided in the Methods section.

### Objectives and aims

The aim of this study was to evaluate the learning effect of an anatomical stereoscopic 3D AR model of the lower leg among medical undergraduates when compared to a monoscopic 3D desktop model and 2D anatomical atlas. The secondary objective was to evaluate whether VSA would modify the observed learning effect. Additionally, the study aimed to evaluate the student's experience of learning anatomy in AR. The authors hypothesized that the stereoscopic 3D AR model is more effective in improving anatomical knowledge than the monoscopic 3D desktop model and the 2D anatomical atlas, and that students with lower levels of VSA benefit most from the stereoscopic 3D view of the model.

## MATERIALS AND METHODS

### Study design

A double-center randomized controlled trial was conducted at the Leiden University Medical Center (LUMC) and the Erasmus University Medical Center Rotterdam (EMC), the Netherlands in the spring of 2018 (Figure 1). The study protocol was approved by the Institutional Review Board at the Leiden University (registration no. CEP17-1215/420). Participation was voluntary and written consent was obtained from all participants.

### Study population

Participants were a volunteer sample of first- and second-year undergraduate students of Medicine and Biomedical Sciences at the LUMC and EMC and were recruited through flyers and announcements during the lectures. The study took place prior to the anatomy courses on the musculoskeletal system of the limbs, ensuring limited knowledge of the lower limb anatomy among all participants. Students who had already taken part in this course were excluded. The baseline knowledge was not assessed to avoid extra burden for students and possible influence on learning during the intervention and the performance on the post-test.[32] Participation in the study did not interfere with the curriculum and the assessment results did not affect student's academic grades. Participants received a compensation of fifteen euros at the completion of the experimental session.

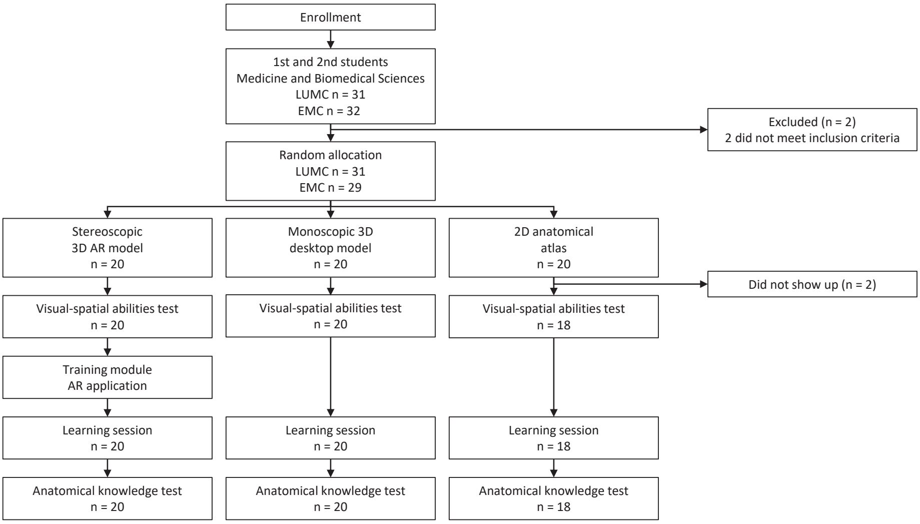


Figure 1. Flowchart of study design. LUMC, Leiden University Medical Center; EMC, Erasmus Medical Center Rotterdam; n, number of participants; AR, Augmented Reality; 3D, three-dimensional; 2D, two-dimensional.

## Randomization

Participants who consented to participate were randomly allocated to either the (1) stereoscopic 3D AR model group, (2) monoscopic 3D desktop model group or (3) 2D anatomical atlas group. Students were assigned an identification number, and these were randomly allocated to the three groups using an Excel Random Group Generator. Blinding of participants was impossible since the intervention was apparent to the students.

## Educational interventions

For the purpose of this study an AR application *DynamicAnatomy* (https://www.microsoft.com/en-us/p/dynamicanatomy/9nwlj4qq053p?SilentAuth=1&activetab=pivot:overviewtab) for Microsoft Hololens® (Version 1, Microsoft, Redmond, Washington, USA) was developed at the Department of Anatomy and Embryology at Leiden University Medical Center and the Centre for Innovation of Leiden University (www.mr4education.com). The application represented a dynamic and fully interactive stereoscopic 3D model of the lower leg. The model was presented as a three-dimensional virtual object in the physical space (*Supplementary material 2*). The Hololens glasses are transparent which enabled participants to stereoscopically interact with the model without losing sense of their own physical environment. A unique feature included an object centered view, i.e., dynamic exploration, which enabled participants to walk around the 3D model and explore it from all possible angles. Participants navigated through the user interface and selected desirable functions by making specific hand gestures or giving a voice command. Active interaction included size adjustments, showing or hiding structures by group or individually, visual and auditory feedback on structures and anatomical layers, and animation of the ankle movements.

With the gaze function switched on, the text of the anatomical descriptions appeared next to the highlighted structure. The anatomical layers included musculoskeletal, connective tissue, and neuro-vascular systems. During this experiment, study participants focused on the musculoskeletal system. Prior to the experiment, participants completed a 10-minutes training module, without anatomical content, to get familiar with the use of the application and device.

For the intended comparison, a Windows desktop application was developed with all the features of *DynamicAnatomy*. The desktop application included the identical anatomical model of the lower limb which was now displayed monoscopically on a 2D computer screen. The model could be rotated along the Y axis in both directions with a slide-bar using a computer mouse *(Supplementary material 3)*. All other features such as voice control, auditory feedback, and scaling were unchanged (Table 1).

Table 1. An overview of available features of the three educational interventions.

| Feature | Stereoscopic 3D AR model | Monoscopic 3D desktop model | 2D anatomical atlas |
|---|---|---|---|
| Stereopsis | + | - | - |
| Dynamic exploration | + | - | - |
| Active user interaction | + | + | - |
| Animation of the ankle | + | + | - |

+ , a feature is present; -, a feature is not present; AR, augmented reality; 3D, three-dimensional; 2D, two-dimensional.

In the 2D anatomical atlas group, study material included selected handouts from an anatomy atlas[33] and an anatomy textbook[34] covering anatomy of the musculoskeletal system. The selection consisted primarily of 2D images of bones and muscles of the lower leg and ankle movements with short descriptions. Each handout included an index for the ease of navigation. In all groups the anatomical descriptions were limited to the names of the structures. No additional textual descriptions were provided.

### Learning objectives and instructional activities

Participants received a handout with a description of the learning goals (identical for each group) and instructions for the learning session (specific for their group). Both were developed based on the constructive alignment theory to ensure the alignment between the intended learning outcomes, instructional activities and knowledge assessment *(Supplementary material 4A, 4B).*[35] The learning goals were formulated and organized according to Bloom's Taxonomy of Learning Objectives.[36] An independent expert outside of the anatomy verified the alignment between the learning goals and the assessment according to the constructive alignment theory and Bloom's Taxonomy of Learning Objectives. Learning goals included memorization of the names of bones and muscles (factual knowledge), understanding the function of the muscles based on their origin and insertion (functional knowledge), and location and organization of these structures in relation to each other (spatial knowledge). Students were free to follow the provided instructions or to choose their own way of achieving the learning goals. Duration of the learning session was 45 minutes.

### Assessment of VSA

VSA were assessed prior to the start of the learning session. Mental visualization and rotation, as the main components of VSA, were assessed by the 24-item Mental Rotation Test (MRT), previously validated by Vandenberg and Kuse[37] (1978) and redrawn by Peters and colleagues[38] (*Supplementary material 5A*). This psychometric test is being widely used in the assessment of VSA and has repeatedly shown its positive association with anatomy learning and assessment.[39,40] The post-hoc level of internal consistency (Cronbach's alpha) of the MRT in this study was 0.88. Mental visualization and transformation, as other components of VSA, were measured by the 10-item Paper Folding Test (PFT), previously validated by Ekstrom and colleagues[41] *(Supplementary material 5B).* The post-hoc level of internal consistency (Cronbach's alpha) of the PFT in this study was 0.76. Additionally, mechanical reasoning was measured by a standardized 12-item Mechanical Reasoning Test (MR), developed for this experiment at the Department of Neuropsychology *(Supplementary material 5C).* The post-hoc level of internal consistency (Cronbach's alpha) of the MR test in this study was 0.76. The duration of the assessment was three minutes for each test. After three minutes all students were instructed to collectively move on to the next test even if they did not finish all the items.

### Anatomy knowledge assessment

The learning effect was evaluated by a 30-item knowledge test. The test consisted of a combination of twenty extended matching questions and ten open-ended questions. The knowledge was assessed in the factual (i.e., memorization/identification of the names of bones and muscles), functional (i.e., understanding the function of the muscles based on their course, origin and insertion) and spatial (i.e., location and organization of structures in relation to each other) knowledge domains (*Supplementary material 6*). Content validation was performed by two experts in the field of anatomy and plastic and reconstructive surgery. The test was then piloted among twelve medical students for item clarity. The post-hoc calculated level of internal consistency (Cronbach's alpha) was 0.78. The duration of the assessment was 30 minutes.

### Evaluation of learning experience

Participants' learning experience was evaluated by a standardized self-reported questionnaire. The evaluation included items on study time, perceived representativeness of the test questions, perceived knowledge gain, usability of and satisfaction with the provided study materials. Response options ranged from "very dissatisfied" (1 point) to "very satisfied" (5 points) on a five-point Likert scale.

### Statistical analysis

Participant's baseline characteristics were summarized using descriptive statistics. The differences in baseline measurements were assessed with a one-way ANOVA for differences in means and $X^2$ test for differences in proportions. The normal distribution was assessed with Shapiro Wilk Test of Normality in combination with the Normal Q-Q Plots. The differences in mean percentages of correct answers on the anatomy knowledge test between groups were assessed with one-way ANOVA including mean percentages of correct answers as a dependent variable and intervention group as a fixed factor. In case of a significant difference, a post-hoc Bonferroni test was performed to identify the pairs of means that differ. The obtained p values were adjusted for multiple comparisons with a Bonferroni correction (*P* value*k). The results were stratified by MRT, PFT and MR test scores to evaluate possible aptitude-treatment interaction between VSA and type of intervention. Additionally, a ANCOVA was performed to evaluate the interaction in a linear regression analysis. Anatomy knowledge test score was included as a dependent variable, intervention group as a fixed factor, VSA test score as a covariate, and 'VSA test score' x 'intervention group' as in interaction term. The effect size (Cohen's d) of the differences in anatomy knowledge test scores between groups was calculated using the mean scores and standard deviations of two groups.[42] 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

A total of sixty participants were included in the study. Two participants allocated to the 2D anatomical atlas group did not show up for the experiment. The 2D anatomical atlas group, therefore, consisted of 18 participants. Participants were not aware of their allocation to one of the three groups in advance but were informed prior to the start of the experiment. Table 2 shows the baseline characteristics of the 58 participants.

Table 2. Baseline characteristics of the included participants.

| | Stereoscopic 3D AR model | Monoscopic 3D desktop model | 2D anatomical atlas | p value |
|---|---|---|---|---|
| | n = 20 | n = 20 | n = 18 | |
| Sex, n (%) | | | | |
| Male | 8 (40.0) | 6 (35.0) | 7 (39.0) | .773 |
| Female | 12 (60.0) | 13 (65.0) | 11 (61.0) | |
| Age, mean ± SD, y | 18.5 ± 0.8 | 18.7 ± 1.0 | 18.7 ± 0.8 | .720 |
| Medical center, n (%) | | | | |
| Leiden University MC | 10 (50.0) | 11 (55.0) | 10 (55.6) | .929 |
| Erasmus University MC Rotterdam | 10 (50.0) | 9 (45.0) | 8 (44.4) | |
| Study, n (%) | | | | |
| Medicine | 17 (85.0) | 16 (80.0) | 14 (77.8) | .842 |
| Biomedical sciences | 3 (15.0) | 4 (20.0) | 4 (22.2) | |
| Study year, n (%) | | | | |
| 1st year | 17 (85.0) | 18 (90.0) | 16 (88.9) | .879 |
| 2nd year | 3 (15.0) | 2 (10.0) | 2 (11.1) | |
| Visual-spatial abilities score, mean ± SD | | | | |
| Mental Rotation Test | 7.1 ± 2.9 | 6.0 ± 2.4 | 8.4 ± 2 .1 | .090 |
| Paper Folding Test | 6.2 ± 1.8 | 6.5 ± 2.6 | 7.6 ± 2.2 | .104 |
| Mechanical Reasoning Test | 9.3 ± 2.6 | 9.3 ± 2.1 | 9.3 ± 3.2 | .990 |

$p < .05$ with a Bonferroni correction for multiple comparison is considered significant. Minimal and maximal scores range between 0-24 for the Mental Rotation Test, 0-10 for the Paper Folding Test and 0-12 for the Mechanical Reasoning Test. n, number of participants; AR, Augmented Reality; 3D, three-dimensional; 2D, two-dimensional; SD, standard deviation; y, years; MC, medical center.

### Overall scores on anatomy knowledge assessment

The scores are presented as mean percentages of correct answers. As shown in Figure 2, the stereoscopic 3D AR group (47.8 %, SD ± 9.8) performed equally well on the knowledge test as the monoscopic 3D desktop group (38.5 %, SD ± 14.3; F(2,54) = 4.79; p = .081) and the 2D anatomical atlas group (50.9 %, SD ± 13.8; F(2,54) = 4.79; p = 1.00). The 2D anatomical atlas group, however, outperformed the monoscopic 3D desktop group (F(2,54) = 4.79; p = .042).

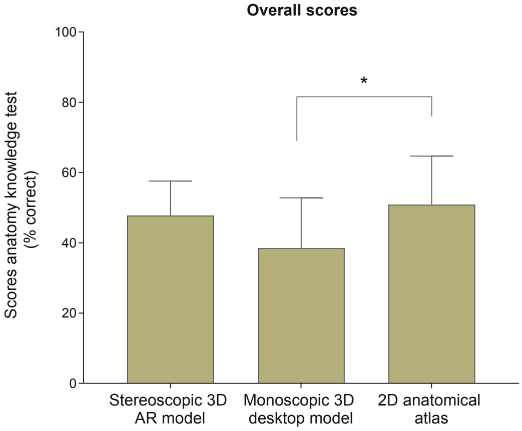


Figure 2. Differences in overall mean percentages correct answers on the anatomy knowledge test between three educational interventions. a, p < .05 analysis of variance with a Bonferroni correction for multiple comparison. MRT, Mental Rotation Test; AR, augmented reality; 3D, three-dimensional; 2D, two-dimensional.

### Stratified by VSA scores

When total scores on the anatomy knowledge test were stratified by Mental Rotation Test (MRT), Paper Folding Test (PFT) and Mechanical Reasoning (MR) test scores, only the MRT scores did significantly impact the outcomes in all three conditions. Students who scored below the mean were assigned to the MRT-low group (n=31) and students who scored above the mean were assigned to the MRT-high group (n=26). As shown in Figure 3, the MRT-high group performed equally well in each of the three intervention groups (F(2,23) = 0.83, p = .448). However, among MRT-low participants significant differences were found between groups. The stereoscopic 3D AR group (49.2 %, SD ± 9.5) significantly outperformed the monoscopic 3D desktop group (33.4 %, SD ± 11.5; F(2,28) = 6.59, p = .015, Cohen's *d* = 1.54), and performed equally well as the 2D anatomical atlas group (46.4 %, SD ± 14.5; F(2,28) = 6.59, p = .990, Cohen's *d* = 0.24). Although, students achieved higher scores in the 2D anatomical atlas group than in the monoscopic 3D desktop group with a moderate effect size (Cohen's *d* = 1.00), the observed difference was not significant (p

= .080). The MRT-low group performed significantly worse than the MRT-high students in the monoscopic 3D desktop group (33.4 %, SD ± 11.5 vs. 49.7 %, SD ± 13.9; p = .015, Cohen's *d* = -1.3) However, they performed equally well in the stereoscopic 3D AR and 2D anatomical atlas groups.

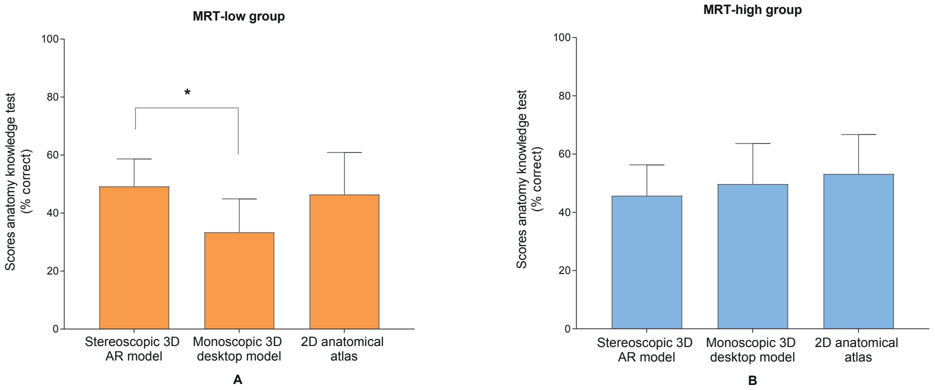


Figure 3. Differences in overall mean percentages correct answers on the anatomy knowledge test between three educational interventions stratified by Mental Rotation Test scores. A, Students who scored below the mean were assigned to the MRT-low group (n = 31) and B, students who scored above the mean were assigned to the MRT-high group (n = 26). $^a$ p < .05 analysis of variance with a Bonferroni correction for multiple comparison. MRT, Mental Rotation Test; AR, augmented reality; 3D, three-dimensional; 2D, two-dimensional.

The observed differences strongly indicate an aptitude-treatment effect caused by VSA. This phenomenon occurs when the effect of an intervention is different in groups of subjects with different characteristics. Therefore, the observed interaction between the MRT scores and the intervention groups was additionally checked in a linear regression analysis. The interaction term 'MRT score' x 'intervention group' showed a marginal trend towards significance (F(2) = 3.04; p = .05). Including PFT and MR test scores as a covariate and an interaction term did not have any significant impact on the outcomes.

## Evaluation of learning experience

As shown in Table 3, participants in the stereoscopic 3D AR group enjoyed the learning session more than the participants in other two groups (4.8 ± 0.4 vs. 3.4 ± 0.8 vs. 2.4 ± 0.9; F(2,54) = 50.3, p = .003). Participants found the application easy and intuitive to use and would recommend it to their fellow students. In all three groups participants reported that their knowledge about anatomy of the lower leg was improved (4.3 ± 0.6 vs. 4.1 ± 0.9 vs. 4.1 ± 0.8; F(2,54) = 0.6, p = .574).

Table 3. Differences in students' learning experience between three educational interventions.

| | Stereoscopic 3D AR group | Monoscopic 3D desktop model | 2D anatomical atlas | p value |
|---|---|---|---|---|
| | n = 20 | n = 20 | n = 18 | |
| The study time was long enough to study the required number of anatomical structures | 3.0 ± 0.9 | 2.3 ± 0.7 | 2.6 ± 0.9 | .192 |
| The questions in anatomy test were representative for the studied material | 3.8 ± 0.6 | 3.8 ± 0.6 | 3.7 ± 0.7 | .709 |
| I enjoyed studying with … | 4.8 ± 0.4$^{\S}$ | 3.4 ± 0.8$^{\S}$ | 2.4 ± 0.9$^{\S}$ | .003$^{*}$ |
| Learning material was easy to use | 4.3 ± 0.6$^{ǂ}$ | 3.4 ± 0.9$^{ǂ}$ | 3.0 ± 0.8$^{ǂ}$ | .009$^{*}$ |
| My knowledge about anatomy of the lower leg is improved after studying with … | 4.3 ± 0.6 | 4.1 ± 0.9 | 4.1 ± 0.8 | .574 |
| I would recommend studying with … to my fellow students | 4.6 ± 0.5$^{\S}$ | 3.7 ± 0.8$^{\S}$ | 2.4 ± 0.9$^{\S}$ | .003$^{*}$ |

Response options on a 5-point Likert scale ranged from 1 = very dissatisfied to 5 = very satisfied on a 5-point Likert scale. Average scores are expressed in means (± SD). * p < .05 analysis of variance with a Bonferroni correction for multiple comparison; § significant difference between all the three groups. ǂ significant difference between (1) Stereoscopic 3D AR model and monoscopic 3D desktop model group; (2) Stereoscopic 3D AR model and 2D anatomical atlas group. n, number of participants; AR, Augmented Reality; 3D, three-dimensional; 2D, two-dimensional; SD, standard deviation.

# DISCUSSION

This study aimed to investigate the educational effectiveness of learning with stereoscopic AR visualization technology and to evaluate whether VSA would modify the learning effect.

Firstly, the observed aptitude-treatment interaction caused by VSA needs to be addressed in more depth. The results showed significant differences in learning effect upon interventions using 2D and 3D learning materials among participants with lower and higher VSA scores as measured by the MRT. These differences were detectible only after stratification of the overall results pointing towards an aptitude-treatment interaction, also referred to as 'effect measure modification'.[27, 43,44] This phenomenon occurs when the effect of an intervention is different in groups of subjects with different characteristics and is different from the effect of a confounder. In current analyses, when VSA were treated only as a confounder, in the absence of stratification, the differences between monoscopic and stereoscopic conditions for different levels of VSA were not evident. This means that an adjustment for this confounder by the study design (e.g., randomization) or statistical analysis (e.g., including it only as a co-variate in the regression analysis), will still not be sufficient, and the results can still be misleading.

Secondly, the monoscopic 3D desktop model group only showed a lower learning effect in the MRT-low group. These findings are supported by previous research in the effectiveness of monoscopic 3D visualization technologies with disadvantages for students with low VSA.[11-15] It has been hypothesized that three-dimensional objects are memorized as key view based two-dimensional images.[13,45] Viewing an unfamiliar 3D object from multiple angles, could therefore lead to an increase in extraneous cognitive load.[16,17,46] The beneficial effect of stereoscopic visualization of a 3D object could be explained by the fact that mental representations depend on the nature of the input.[20,24, 25] In that case, mental representations do not primarily consist of key view based 2D images, but they might also include spatial information. This is further supported by the evidence that disparity processing occurs in different visual pathways of the human brain.[26] This means, that while a monoscopic 3D desktop view and 2D anatomical atlas images would stimulate key view based 2D mental images, a stereoscopic 3D model would stimulate structural 3D mental representations. Stereopsis might then avoid the increase in extraneous cognitive load and therefore facilitate a better comprehension of 3D anatomy in students with lower levels of VSA.

As dynamic exploration was the second distinguishing feature of the stereoscopic 3D AR model, it may also have contributed to the positive learning effect. Being able to walk around the model with its own reference point can create an additional sense of depth. On the other hand, the object centered view is different from the egocentric view where the user moves the objects in their field with virtual tools, as was the case in the monoscopic 3D desktop group. The egocentric control can affect visual-spatial skills where the hands are involved in imagining the rotation of objects. Future research is needed to evaluate how these different types of view in a 3D environment affect spatial processing during learning. This should be performed in an identical environment using the same medium, configuration and presentation.[27] This eliminates all possible confounding effects of additional features such as hand gestures, that can vary between different types of media.

Thirdly, participants in the 2D anatomical atlas group achieved anatomy knowledge test scores similar to those in the stereoscopic 3D AR model group. This unexpected effect can be hypothetically explained by several reasons. One is the 2D nature of the paper-pencil assessment which in fact was more aligned with the studied material in the 2D anatomical atlas group. In a recent study on the effectiveness of a monoscopic 3D visualization technology versus the use of prosected cadaveric specimens, students have performed best on the identification questions aligned with the respective study materials.[47,48] A similar effect has been reported by Henssen and colleagues[49] with the use of cross-sections. Therefore, participants in the 2D anatomical atlas group could have had an advantage over participants in the other two groups. More insight can be gained by future studies that include a combination of assessment methods aligning with each of the interventions.

Another explanation is of a more theoretical nature, namely the unfamiliarity with a new type of 3D visualization technology and the meta-representational competence of students as part of their spatial intelligence. Hegarty has described this competence as the ability to choose the optimal external representation for a task, and effective use of novel external representations, such as interactive visualizations.[50] In their research, novice Navy weather forecasters tended to choose less effective interactive visualization than experts by adding unnecessary visual information to a display in order to interpret a weather forecast.[51] In the current study, relevant 2D images were selected form the anatomical atlas which made it easy for students to identify quickly the useful images. In the intervention group, however, students had to rely on their own choices of visual representations. In an interactive 3D environment, students with lower visual-spatial abilities could therefore be less effective in choosing the right representations of anatomical structures to learn from (e.g., exploring an anatomical structure in the presence of all other structures and/or menu options versus isolating a structure from all other anatomical layers and restricting the user interface to a minimal amount of visual information). Additionally, students with lower VSA tend to use the interactive presentations less effectively. These students for example had difficulties in rotating a digital 3D anatomical structure to a specified view.[50,52] However, with the aid of orientation references, students have been able to successfully manipulate and learn from the virtual model. The tendency to choose a less effective strategy by low performing students has recently been demonstrated by Roach and colleagues in performing a mental rotation task.[53-55] Students with high VSA had a distinct eye-movement pattern in solving mental rotation tasks than low performing students.[53] When low performing students had been instructed by a visual guidance protocol that was based on the eye-movement pattern of high performing students, they had significantly improved in solving the mental rotation tasks.[55] For the reasons stated above, these individual differences can potentially affect the learning strategies of students and are of great interest for further investigation.

## Future directions

The findings have implications for both research and education. The modifying effect of VSA should be taken into account when designing new research and analysis strategies, especially in the field of 3D technologies. For educational purposes, stereoscopic 3D AR models have a great potential to be effectively used in small-group teaching settings to stimulate active learning and peer-to-peer interaction by studying a synchronized anatomical 3D models. In addition to traditional ways of teaching, this new teaching tool can be used in the context of personalized learning to meet the students' individual learning needs. Especially, the combination of stereoscopic 3D models and 2D anatomical atlas is worth further research. A possible synergic learning effect would be desirable since the level of anatomical knowledge among medical students still remain insufficient.[1–5] When designing new VR and AR environments one should carefully align the learning environment with the (learning) goals, e.g., VR is better suited for individual learning experiences, whereas AR has many advantages for collaborative and embodied learning.

## Limitations

There are some methodological limitations in this study. Firstly, due to the limited availability of hardware, the study was restricted to a maximum of twenty participants in each group. In addition, no distributional data on anatomy knowledge assessment was available beforehand. Therefore, an a priori sample size calculation could not be performed. Only for this reason, a post hoc power analysis was performed based on the observed effect sizes, which turned out to be sufficient. Second concern was the alignment between study materials and assessment. A different form of assessment that is closer to the clinical practice and in line with the learning method (e.g., cadaveric/specimen or digital 3D assessment) should be considered to assess the acquired anatomical knowledge. If not possible, a combination of assessment methods aligning with each of the interventions should be considered. In addition, a long-term retention test would have been valuable to measure the actual retention of anatomical knowledge. Thirdly, the participants were not tested for their lack of depth perception which could be present in about five percent of the study population.[56] Based on these statistics, 1-2 of the twenty participants in the stereoscopic 3D AR group could have perceived the model monoscopically, which could have unfairly lowered the total group score. Lastly, some of the features that were characteristic for the type of intervention, for example hand gestures in stereoscopic 3D AR group and audio cues in both stereoscopic 3D AR and monoscopic 3D desktop groups, could have introduced bias. To eliminate such differences between groups, it is desirable to conduct research within one level of instructional design when possible. Additionally, this will decrease the chance of Hawthorne effect that can occur when learners tend to learn better or harder with a more popular tool or medium, as it could have been the case in the current study.

# CONCLUSIONS

Three-dimensional anatomical models that can be viewed stereoscopically in AR can help to optimize anatomical knowledge acquisition and knowledge of students with lower levels of VSA. Further research is needed to identify factors that contribute to the positive learning effect and the most effective way of combining this technology with current education.

# REFERENCES

1. 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 2010; 37:954–961.
2. 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.
3. Spielmann PM, Oliver CW. The carpal bones: A basic test of medical students and junior doctors' knowledge of anatomy. Surgeon 2005;3:257–259.
4. Waterston SW, Stewart IJ. Survey of clinicians' attitudes to the anatomical teaching and knowledge of medical students. Clin Anat 2005;18:380–384.
5. Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. How much anatomy is enough? Anat Sci Educ 2008;1:184–188.
6. Pryde FR, Black SM. Anatomy in Scotland: 20 years of change. Scott Med 2005;50:96–98.
7. 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.
8. 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.
8. Klatzky RL, Lederman SJ. Haptic object perception: Spatial dimensionality and relation to vision. Philos Trans R Soc Lond B Biol Sci 2011;366:3097–3105.
9. Bergman EM, de Bruin AB, Herrler A, Verhrijen IW, Scherpbier, van der Vleuten CP. Students' perceptions of anatomy across the undergraduate problem-based learning medical curriculum: A phenomenographical study. BMC Med Educ 2013;13:152–162.
9. Reid S, Shapiro L, Louw G. How haptics and drawing enhance the learning of anatomy. Anat Sci Educ 2018;12:164–172.
10. 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.
11. Garg A, Norman GR, Spero L, Maheshwari P. Do virtual computer models hinder anatomy learning? Acad Med 1999;74:S87–S89.
12. Garg A, Norman G, Spero L, Taylor I. Learning anatomy: Do new computer models improve spatial understanding? Med Teach 1999;21:519–522.
13. 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.
14. Levinson AJ, Weaver B, Garside S, McGinn H, Norman GR. Virtual reality and brain anatomy: a randomised trial of e-learning instructional designs. Med Educ 2007;41:495–501.
15. Naaz F. Learning from graphically integrated 2D and 3D representations improves retention of neuroanatomy. University of Louisville: Louisville, KY. Doctorate of Philosophy Dissertation. 2002. 76 p.
16. Huk T. Who benefits from learning with 3D models? The case of spatial ability. J Comput Assist Learn 2006;22:392–404.
17. 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;6:211–215.
18. 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.

19. Luursema JM, Verwey WB, Kommers PA, Geelkerken RH, Vos HJ. Optimizing conditions for computer-assisted anatomical learning. Interact Comput 2006;18:1123–1138.
20. Luursema JM, Verwey WB, Kommers PA, Annema JH. The role of stereopsis in virtual anatomical learning. Interact Comput 2008;20:455–460.
21. Luursema JM, Vorstenbosch M, Kooloos J. Stereopsis, visuospatial ability, and virtual reality in anatomy learning. Anat Res Int 2017:1493135.
22. Hackett M, Proctor M. The effect of autostereoscopic holograms on anatomical knowledge: A randomized trial. Med Educ 2018;52:1147–1155.
23. 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.
24. Jolicoeur P, Milliken B. Identification of disoriented objects: effects of context of prior presentation. J Exp Psychol Learn Mem Cogn 1989;15:200–210.
25. Kourtzi Z, Erb M, Grodd W, Bulthoff HH. Representation of the perceived 3-D object shape in the human lateral occipital complex. Cereb Cortex 2003;13:911–920.
26. Verhoef B-E, Vogels R, Janssen P. Binocular depth processing in the ventral visual pathway. Philos Trans R Soc Lond B Biol Sci 2016;371:20150259.
27. Cook DA. The research we still are not doing: An agenda for the study of computer-based learning. Acad Med 2005;80:541–548.
27. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017;10:549–559.
28. Kuehn BM. Virtual and augmented reality put a twist on medical education. JAMA 2018;319:756–758.
29. Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load. Anat Sci Educ 2016; 9:411–421.
30. Barmaki R, Yu K, Pearlman R, Shingles R, Bork F, Osgood GM, Navab N. Enhancement of anatomical education using augmented reality: An empirical study of body painting. Anat Sci Educ 2019;12:599–609.
31. Sugiura A, Kitama T, Toyoura M, Mao X. The use of augmented reality technology in medical specimen museum tours. Anat Sci Educ 2019;12:561–571.
32. Cook DA, Beckman TJ. Reflections on experimental research in medical education. Adv Health Sci Educ Theory Pract 2010;15:455–464.
32. McJunkin JL, Jiramongkolchai P, Chung W, Southworth M, Durakovic N, Buchman CA, Silva JR. Development of a mixed reality platform for lateral skull base anatomy. Otol Neurotol 2018;39:e1137–e1142.
33. Putz R, Pabst R. Sobotta Atlas of Human Anatomy. Part 2. 3rd Ed. Houten, The Netherlands: Bohn Stafleu van Loghum. 2006; 399 p.
34. Moore KL, Dalley AF, Agur AM. Clinically Oriented Anatomy. 7th Ed. Philadelphia, PA: Lippincott Williams & Wilkins. 2013;1168 p.
35. Biggs J. Enhancing teaching through constructive alignment. High Educ 1996;32:347–364.
36. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. 1st Ed. New York, NY: David McKay Company. 1956. 207 p.
37. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills 1978;47:599–604.

38. 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.
39. Guillot A, Champely S, Batier C, Thiriet P, Collet C. Relationship between spatial abilities, mental rotation and functional anatomy learning. Adv Health Sci Educ Theory Pract 2007;12:491–507.
40. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ 2017;10:235–241.
41. Ekstrom RB, French J, Harman HH, Dermen D.Kit of Factor-Referenced Cognitive Tests. Princeton, NJ: Educational Testing Service. 1976. 314 p.

42 Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Ed. Hillsdale, NJ: Lawrence Earlbaum Associates. 1988. 400 p.

43. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd Ed. Philadelphia, PA: Lippincott Williams and Wilkins. 2008. 758 p.
44. Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandenbroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. Clin Epidemiol 2017;9:331–338.
45. Bulthoff HH, Edelman SY, Tarr MJ. How are three-dimensional objects represented in the brain? Cereb Cortex 1995;5:247–260.
46. Mayer RE. Cognitive theory of multimedia learning. In: Mayer RE (Editor). Multimedia Learning. 2nd Ed. Santa Barbara, CA: Cambridge University Press. 2014. p 43–47.
47. Mitrousias V, Karachalios TS, Varitimidis SE, Natsis K, Arvanitis DL, Zibis AH. Anatomy learning from prosected cadaveric specimens versus plastic models: A comparative study of upper limb anatomy. Anat Sci Educ 2020;13:436-444.
48. Mitrousias V, Varitimidis SE, Hantes ME, Malizos KN, Arvanitis DL, Zibis AH. 2018. Anatomy learning from prosected cadaveric specimens versus three-dimensional software: A comparative study of upper limb anatomy. Ann Anat 218:156–1564.
49. Henssen DJ, van den Heuvel L, De Jong G, Vorstenbosch MA, van Cappellen van Walsum AM, Van den Hurk MM, Kooloos JG, Bartels RH. 2020. Neuroanatomy learning: Augmented reality vs. cross-sections. Anat Sci Educ (in pres; doi 10.1002/ase.1912).
50. Hegarty M. Chapter 7 - Components of spatial intelligence. Psychol Learn Motiv 2010;**52**:265–297.
51. Smallman HS, Hegarty M. Expertise, spatial ability and intuition in the use of complex visual displays. In: Proceedings of the 51st Annual Meeting of the Human Factors and Ergonomics Society (HFES 2007); Baltimore, MD, 2007 October 1-5. p 2000–2004. Human Factors and Ergonomics Society: Santa Monica, CA.
52. Stull AT, Hegarty M, Mayer RE. Orientation references: Getting a handle on spatial learning. J Educ Psychol 2009;101:803–816.
53. 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.
54. 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.
55. 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.
56. Mather G. Foundations of Perception. 1st Ed. Hove, UK: Psychology Press. 2006. 400 p.

# 4

# The effect of binocular disparity on learning anatomy with stereoscopic augmented reality visualization:

## a double-center randomized controlled trial

Katerina Bogomolova, Marc A.T.M. Vorstenbosch, Inssaf El Messaoudi,
Micha Holla, Steven E. R. Hovius, Jos A. van der Hage, Beerend P. Hierck

*Accepted for publication in Anatomical Sciences Education*

# ABSTRACT

## Background

Binocular disparity provides one of the important depth cues within stereoscopic three-dimensional (3D) visualization technology. However, there is limited research on its exact effect on learning within 3D augmented reality (AR) environment. This study evaluated the effect of binocular disparity on acquisition of anatomical knowledge and perceived cognitive load in relation to visual-spatial abilities (VSA).

## Methods

In a double-center randomized controlled trial, first-year (bio)medical undergraduates studied lower extremity anatomy in an interactive 3D AR environment either with a stereoscopic 3D view ($n = 32$) or monoscopic 3D view ($n = 34$). VSA were tested with a Mental Rotation Test (MRT). Anatomical knowledge was assessed by a validated 30-item written test and 30-item specimen test. Cognitive load was measured by the NASA-TLX questionnaire.

## Results

Students in the stereoscopic 3D and monoscopic 3D groups performed equally well in terms of percentage correct answers (written test: 47.9 ± 15.8 vs 49.1 ± 18.3; $p = .635$; specimen test: 43.0 ± 17.9 vs 46.3 ± 15.1; $p = .429$), and perceived cognitive load scores (6.2 ± 1.0 vs 6.2 ± 1.3; $p = .992$). Regardless of intervention, VSA were positively associated with the specimen test scores ($\eta^2 = 0.13$, $p = .003$), perceived representativeness of the anatomy test questions ($p = .010$) and subjective improvement in anatomy knowledge ($p < .001$).

## Conclusions

In conclusion, binocular disparity does not improve learning anatomy. Motion parallax should be considered as another important depth cue that contributes to depth perception during learning in a stereoscopic 3D AR environment.

# INTRODUCTION

Three-dimensional visualization technology (3DVT) has a great potential to contribute to a better learning and understanding of anatomy in medical education. [1,2] Its contribution is becoming necessary in times of decreased teaching hours of anatomy and exposure to traditional teaching methods, such as cadaveric dissections.[3-5] Additionally, anatomical knowledge is reported to be insufficient among medical students and junior doctors, who still experience difficulties in translating the acquired knowledge into clinical practice.[6-9] However, to know *that* 3DVT can be highly effective, is currently not enough. There is a need to know *how* this technology works to be able to implement it in everyone's unique educational setting.[10]

## Stereoscopic versus monoscopic three-dimensional visualization technology

In real life, stereoscopic vision is obtained due to positioning of the human eyes in a way that generates two slightly different retinal images of an object, also referred to as binocular disparity.[11] The same effect can be mimicked within 3DVT by presenting a slightly shifted and rotated image to the right and left eye. Stereoscopic vision can be obtained by supportive devices such autostereoscopic displays e.g., Alioscopy 3D Display (Alioscopy, Paris, France), anaglyphic or polarized glasses, or by a head-mounted display e.g., HoloLens™ (Microsoft Corp., Redmond, WA), or Oculus Rift™ (Oculus VR, Menlo Park, CA) and HTC VIVE™ (High Tech Computer Corp., New Taipei City, Taiwan). Hololens™ is used to create interactive augmented reality (AR), also referred to as mixed reality (MR). Oculus Rift™ and HTC VIVE™ are predominantly used to create virtual reality (VR) environments. A binocular vision of the viewer, though, is required to perceive the obtained visual depth. In the absence of stereoscopic vision, 3D effect is mimicked by monocular cues, such as shading, coloring, relative size and motion parallax.[12] The examples of monoscopic 3DVT include 3D anatomical models that can be explored from different angles on a computer, tablet or phone.[13]

Distinction between stereoscopic and monoscopic modalities within 3DVT is essential to make since different processes are involved. Research has shown that recognition of digital 3D objects appears to be greater when objects are presented stereoscopically.[14-17] More importantly, the type of modality can significantly affect learning. Monoscopic 3DVT has been demonstrated to have disadvantages for students with lower VSA.[18-23] The disadvantages are explained by the ability-as-enhance hypothesis within the cognitive load theory.[24,25] Initially, it has been hypothesized that 3D objects are remembered as key view-based 2D images.[20,21,26]

Consequently, when an unfamiliar 3D object is viewed from multiple angles, an increase in cognitive load occurs while generating a proper mental representation of a 3D object. During this process, individuals with higher visual-spatial abilities are able to devote more cognitive resources to building mental connections, while students with low VSA get cognitively overloaded.[20,26] The latter leads to underperformance among students with lower VSA. However, as research has shown, with stereoscopic 3DVT, students with lower VSA are able to achieve comparable levels of performance of students with higher VSA.[28] This can be explained by the fact that the mental 3D representations of the object are already built and provided by the stereoscopic projection and perception. Consequently, mental steps, that are required to build a mental 3D representation, can be skipped while leaving a sufficient amount of cognitive resources. In this way, students with lower VSA are able to allocate these resources to learning.

### The role of stereopsis

In health care, the benefits of stereoscopic visualization within 3D technologies have been recognized for years.[29-33] Development and utilization of stereoscopic 3DVT is still growing, especially in the surgical field. Several examples include preoperative planning and identification of tumor with stereoscopic AR.[31,34,35] Another examples include stereoscopic visualization during minimal invasive surgeries where stereoscopic view of the surgical field would improve spatial understanding and orientation during laparoscopic surgeries.[29,36] Stereoscopic visualization even showed to shorten operative time of laparoscopic gastrectomy by reducing the intracorporeal dissection time.[37]

The beneficial effect of stereopsis on learning anatomy has been recently demonstrated in a comprehensive systematic review and meta-analysis.[38] In the meta-analysis, the comparisons between studies were made within a single level of instructional design, e.g., stereopsis was isolated as the only true manipulated element in the experimental design. The positive effect of stereopsis was demonstrated across different types of 3D technologies combined, predominantly using the VR headsets and 3D shutter glasses for desktop applications. How learning experience is affected by a particular type of stereoscopic 3DVT, remains a topic for further exploration.

### Stereoscopic augmented reality in anatomy education

Stereoscopic AR is a new generation of 3DVT technology that combines stereoscopic visualization of 3D computer-generated objects with the physical environment. The main distinguishing feature from other types of AR is the ability to provide stereoscopic vision, e.g., to perceive the anatomical model in real 3D. Additionally, it provides the ability to walk around the model and explore it form all possible angles without losing the sense of the user's own environment. This view can be obtained with e.g., HoloLens®, a head-mounted display from Microsoft (*Supplementary material 2*). In the previous study, authors evaluated the effectiveness of stereoscopic 3D AR visualization in learning anatomy of the lower leg among medical undergraduates.[23] Learning with a stereoscopic 3D AR model was more effective than learning with a monoscopic 3D desktop model. Interestingly, the observed positive learning effect was only present among students with lower VSA. Stereoscopic vision was hypothesized to be one of the distinguishing features of intervention modality that could explain these differences. However, since the comparisons were made within different levels of instructional design, e.g., stereoscopic vision was not isolated as the only true manipulated element, the actual effect of stereoscopic vision remained unrevealed. A similar study design approach was used by Moro and colleagues who compared the effectiveness of HoloLens with mobile-based AR environment.[39] Although both learning modes were effective in terms of acquired anatomical knowledge, comparisons were still made within different levels of instructional design.

In another recent study of the role of stereopsis in 3DVT, Wainman and colleagues have isolated binocular disparity by covering the non-dominant eye of participants. Authors reported positive effect of stereoscopic vision in VR, but not in AR.[40] Although it was a simple and vivid way of isolating stereopsis, participants in the control group remained aware of their condition which could have influenced the outcomes. Additionally, different effect measures of stereopsis in VR and AR suggest that the type of technology is decisive for the learning effect caused by stereoscopic vision.

### Objectives and aims

Based on considerations described above and lessons learned from previous research, this study aimed to evaluate the role of binocular disparity in a stereoscopic 3D AR environment within a single level of instructional design. Therefore, the primary objective was to evaluate whether learning with a stereoscopic view of a 3D anatomical model of the lower extremity was more effective than learning with a monoscopic 3D view of the same model among medical undergraduates. The secondary objectives were to compare the perceived cognitive load among groups, and to evaluate whether VSA would modify the outcomes.

Authors hypothesized that learning within a stereoscopic 3D AR environment would be more effective than learning within a monoscopic 3D AR environment. Authors also hypothesized that the perceived cognitive load in the stereoscopic 3D view group would be lower than in the monoscopic 3D view group, and that the students with lower VSA would benefit most from the stereoscopic 3D view of the model.

# MATERIALS AND METHODS

## Study design

A single-blinded double-center randomized controlled trial was conducted at the Leiden University Medical Center (LUMC) and the Radboudumc University Medical Center (Radboudumc), the Netherlands. The study was conducted within a single level of instructional design, e.g., isolating binocular disparity as the only true manipulated element. (Figure 1). The study was approved by the Netherlands Association for Medical Education (NVMO) Ethical Review Board (NERB case number: 2019.5.8).

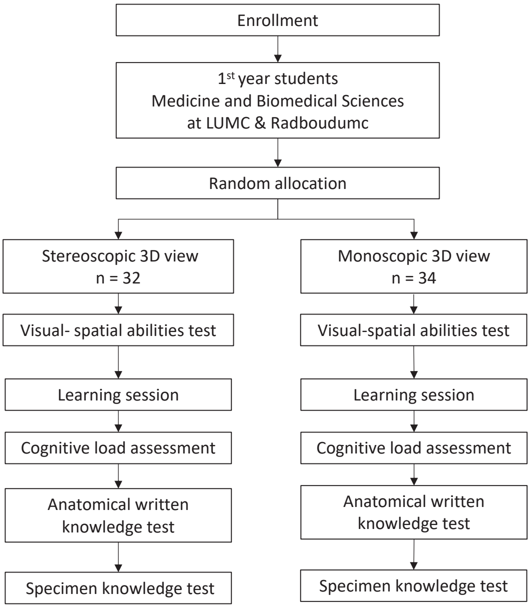


Figure 1. Flowchart of study design. LUMC, Leiden University Medical Center; Radboudumc, Radboud University Medical Center; 3D, three-dimensional; n, number of participants.

## Study population

Participants were first-year undergraduate students of Medicine and Biomedical Sciences with no prior knowledge of the lower extremity anatomy. The baseline knowledge was not assessed to avoid extra burden for students and possible influence on learning during the intervention and performance on the post-tests. [41] Participation was voluntary and written consent was obtained from all participants. Participation did not interfere with the curriculum and the assessment results did not affect student's academic grades. Participants received a financial compensation at the completion of the experiment.

### Randomization and blinding of participants

Participants were randomly allocated to either stereoscopic 3D view or monoscopic 3D view groups using an Excel Random Group Generator (Microsoft Excel for Office 365 MSO, version 2012). Participants were not aware of the distinction between stereoscopic and monoscopic 3D views and remained blinded to the type of condition during the entire experiment. The intended goal of the study and individual allocation to study arms was clarified and debriefed directly after experiment.

### Educational interventions

An interactive AR application *DynamicAnatomy* for Microsoft HoloLens®, version 1 (Microsoft Corp., Redmond, WA) was developed at the department of Anatomy at Leiden University Medical Center and the Centre for Innovation of Leiden University. The application represented a dynamic and fully interactive stereoscopic 3D model of the lower extremity. Users perceived the 3D model as a virtual object in their physical space without losing the sense of their own physical environment. The object centered view, i.e., dynamic exploration, enabled learners to walk around the model and explore it from all possible angles. Active interaction included size adjustments, showing or hiding anatomical structures by group or individually, visual and auditory feedback on structures and anatomical layers, and animation of the ankle movements. The anatomical layers included musculoskeletal, connective tissue, and neurovascular systems. During this experiment, study participants studied the musculoskeletal system. Prior to the experiment, participants completed a 10-minutes training module (without anatomical content) to get familiar with the use of application and device.

In the intervention group, the 3D model of the lower extremity was presented and perceived stereoscopically as intended by the supportive AR device. In the control group, binocular disparity was eliminated technically by projecting an identical, i.e., non-shifted and non-rotated, image to both eyes. This adjustment resulted in a monoscopic view of the identical 3D anatomical model. Therefore, binocular disparity was isolated as the only true manipulated element in this experimental design. All other features of the AR application described above remained available and identical in both conditions.

### Stereovision of participants

Stereovision of participants was measured by a Random Dot 3 - LEA Symbols® Stereoacuity Test (Vision Assessment Corp., Elk Grove Village, IL) prior to the experiment to identify individuals with absent stereovision.

### Baseline characteristics

Informed consent and baseline questionnaire were administered prior to the start of the experiment.

### Assessment of VSA

VSA were assessed prior to the start of the learning session. Mental visualization and rotation, as the main components of visual-spatial abilities, were assessed by the 24-item Mental Rotation Test (MRT), previously validated by Vandenberg and Kuse and redrawn by Peters and colleagues.[42,43] This psychometric test is being widely used in the assessment of VSA and has repeatedly shown its positive association with anatomy learning and assessment.[44,45] The post-hoc level of internal consistency (Cronbach's alpha) of the MR test in this study was 0.94. Duration of this test was ten minutes without intervals.

### Learning session

Participants received a handout with a description of the learning goals and instructional activities. The development of learning goals and instructions was based on the constructive alignment theory to ensure alignment between the intended learning outcomes, instructional activities and knowledge assessment (*Supplementary material 4A, 4B*).[23] Learning goals were formulated and organized according to Bloom's Taxonomy of Learning Objectives.[46] An independent expert verified the alignment between the learning goals and the assessment according to the constructive alignment theory and Bloom's Taxonomy of Learning Objectives. Learning goals included memorization of the names of bones and muscles, understanding the function of muscles based on their origin and insertion, and location and organization of these structures in relation to each other. Duration of the learning session was 45 minutes.

### Cognitive load assessment

Cognitive load was measured by the validated NASA-TLX questionnaire immediately after the session.[47] The NASA-TLX questionnaire is a subjective, multidimensional assessment instrument for perceived workload of task, in this case the workload required to study the anatomy of lower extremity. The items included mental demand, physical demand, temporal demand, performance, effort, and frustration level. Response options ranged from low (0 point) to high (10 points). The total score was calculated according to the prescriptions of the questionnaire and ranged also between 0 and 10 points.

### Written anatomy knowledge test

A previously validated 30-item knowledge test consisted of a combination of 20 extended matching questions and 10 open-ended questions (*Supplementary material 6*).[23] Anatomical knowledge was assessed in the factual (i.e., memorization/identification of the names of bones and muscles), functional (i.e., understanding the function of the muscles based on their course, origin and insertion) and spatial (i.e., location and organization of structures in relation to each other) knowledge domains. Content validation was assessed by two experts in the field of anatomy and plastic and reconstructive surgery. The test was piloted among 12 medical students for item clarity. The level of internal consistency (Cronbach's alpha) was 0.78. Duration of assessment was 30 minutes.

### Specimen knowledge test

Plastinated specimen test covered a total of 30 anatomical structures on 12 specimens distributed over 10 stations (*Supplementary material 7).* Content validation was assessed by one expert in the field of anatomy. Each station included 3-4 structures that were labeled on one or more specimen. Participants were asked to name the labeled structures or to indicate what movement is initiated by a particular structure. The post-hoc level of internal consistency (Cronbach's alpha) of the test was 0.90. Duration of this assessment was 20 minutes with a maximum of two minutes per station.

### Evaluation of learning experience

Participants' learning experience was evaluated by a standardized self-reported questionnaire. The evaluation included items on study time, perceived representativeness of the test questions, perceived knowledge gain, usability of and satisfaction with the provided study materials. Response options ranged from "very dissatisfied" (1 point) to "very satisfied" (5 points) on a five-point Likert scale.

### Statistical analysis

Participant's baseline characteristics were summarized using descriptive statistics. The differences in baseline measurements were assessed with an independent t-test for differences in means and $X^2$ test for differences in proportions. Anatomical knowledge was defined as mean percentage of correct answers on the written knowledge test and the specimen test. Cognitive load was defined as the mean score on the NASA-LTX questionnaire. Differences in outcome measures between groups were assessed with an independent t-test. Additionally, a ANCOVA was performed to measure the effect of the intervention for different levels of visual-spatial abilities by including the interaction term "MRT score x intervention" in the model. MRT score was also included as a covariate to measure its effect on outcomes regardless of intervention. Additional analyses were performed for sex differences. 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

A total of 66 students were included (Table 1). All participants were able to perceive spatial visual depth as measured by the stereoacuity test.

Table 1. Baseline characteristics of included participants.

| | Stereoscopic 3D view | Monoscopic 3D view | p value |
|---|---|---|---|
| | n = 32 | n = 34 | |
| Sex, n (%) | | | |
| Male | 11 (34.4) | 16 (47.1) | .295 |
| Female | 21 (65.6) | 18 (52.9) | |
| Age, mean ±SD, y | 19.2 ± 1.3 | 19.0 ± 1.9 | .754 |
| Medical center, n (%) | | | |
| Leiden University MC | 16 (50.0) | 16 (47.1) | .811 |
| Radboudumc University MC | 16 (50.0) | 18 (52.9) | |
| Study, n (%) | | | |
| Medicine | 30 (93.8) | 33 (97.1) | .519 |
| Biomedical sciences | 2 (6.3) | 1 (2.9) | |
| Videogame, n (%) | | | |
| Never | 23 (71.9) | 22 (64.7) | .397 |
| 0-2 hours a week | 6 (18.8) | 6 (17.6) | |
| 2-10 hours a week | 2 (6.3) | 6 (17.6) | |
| >10 hours a week | 1 (3.1) | 0 (0.0) | |
| AR experience before, n (%) | | | |
| No | 26 (81.3) | 23 (67.6) | .207 |
| Yes | 6 (18.8) | 11 (32.4) | |
| Mental Rotation Test, mean ±SD | 14.1± 5.1 | 15.7 ± 5.3 | .212 |

Minimal and maximal scores range between 0-24 for the Mental Rotation Test, 3D, three-dimensional; AR, Augmented Reality; n, number of participants; SD, standard deviation; y, years; MC, medical center.

As shown in Figure 2, participants in the stereoscopic 3D view group performed equally well as the participants in the monoscopic 3D view group on the written knowledge test (47.9 ± 15.8 vs 49.1 ± 18.3; p = .635). Likewise, no differences were found for each knowledge domain separately (*factual:* 34.1 ± 19.5 vs 34.3 ± 19.0; p = .970; *functional:* 33.4 ± 16.4 vs 31.5 ± 13.7; p = .611; *spatial:* 50.4 ± 15.2 vs 47.3 ± 13.5; p = .384).

Percentages correct answers on the specimen test were not significantly different between groups (43.0 ± 17.9 vs 46.3 ± 15.1; p = .429) (Figure 2).

The observed similarities between groups on the knowledge tests were reflected by the cognitive load scores that were similar in both groups (6.2 ± 1.0 vs 6.2 ± 1.3; p = .992) (Figure 2).

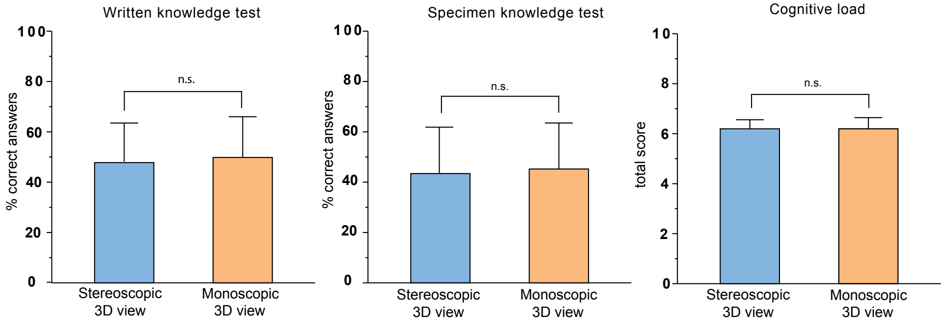


Figure 2. Differences in overall mean percentages correct answers on the written knowledge test, specimen knowledge test and cognitive load test between stereoscopic 3D view (n = 32) and monoscopic 3D view (n = 34) groups. 3D, three-dimensional; n.s., not significant.

As shown in Table 2, there were no significant differences in learning experience between stereoscopic 3D view and monoscopic 3D view groups. All participants enjoyed studying (4.4 ± 0.7 vs 4.3 ± 0.8; p = .492) and reported an improved anatomical knowledge of the lower extremity (4.2 ± 0.9 vs 4.1 ± 0.7; p = .502). Five versus four participants reported the device to be heavy on their nose after a longer period of study time in stereoscopic and monoscopic 3D groups respectively (p = .794). Headache and nausea were reported by one participant in the stereoscopic 3D group.

Table 2. Differences in learning experience between groups.

| | Stereoscopic 3D view | Monoscopic 3D view | p value |
|---|---|---|---|
| | n = 32 | n = 34 | |
| The study time was long enough to study the required number of anatomical structures | 2.4 ± 1.0 | 2.5 ± 1.1 | .500 |
| The questions in anatomy test were representative for the studied material | 3.7 ± 1.0 | 3.4 ± 1.1 | .249 |
| My knowledge about anatomy of the lower leg is improved after studying | 4.2 ± 0.9 | 4.1 ± 0.7 | .502 |
| Learning material was easy to use | 3.5 ± 0.9 | 3.4 ± 0.9 | .378 |
| I enjoyed studying | 4.4 ± 0.7 | 4.3 ± 0.8 | .492 |
| I would recommend studying with … to my fellow students | 4.1 ± 0.8 | 3.7 ± 0.9 | .057 |

Response options on a five-point Likert scale ranged from 1 = very dissatisfied to 5 = very satisfied. Scores are expressed in means ± SD. SD, standard deviation; 3D, three-dimensional; n, number of participants.

### The effect of VSA

In both study groups, mean scores on the written knowledge test and for each knowledge domain separately remained similar for all levels of MRT scores, as measured by the interaction term in the ANCOVA analysis (written knowledge test: F(1,62) = 0.51, p = .393; *factual:* F(3,62) = 0.15, p = .925; *functional:* F(3,62) = 1.04, p = .381; *spatial:* F(3,62) = 0.92, p = .435).

Similar effects were found for the specimen knowledge test (F(1,62) = 0.00, p = .998). However, regardless of intervention, MRT scores were significantly and positively associated with the specimen test scores, as shown in Figure 3 (F(1,62) = 9.37, Partial η2 = 0.13, p = .003).

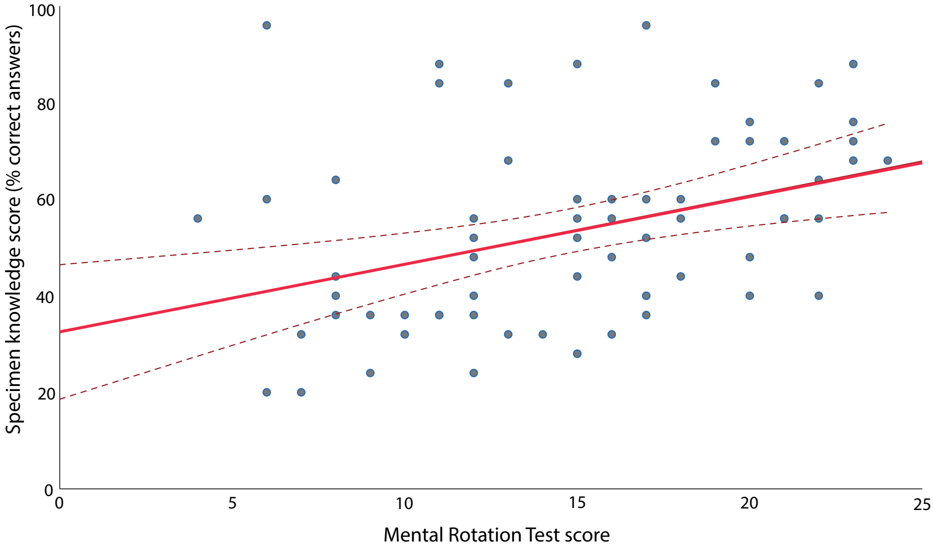


Figure 3. Relationship between Mental Rotation Test (MRT) scores and specimen test scores. A regression analysis graph illustrating a positive association between visual-spatial abilities and specimen test scores. MRT, Mental Rotations Test.

The perceived cognitive load scores remained similar for all levels of VSA in both study groups (F(1,62) = 2.26, p = .138). Regardless of intervention, MRT scores were not associated with the perceived cognitive load scores.

ANCOVA analysis for learning experience revealed that participants in the monoscopic 3D view group found the anatomy test questions significantly less representative for the studied material than participants in the stereoscopic 3D view group. This difference was only present among individuals with lower visual-spatial abilities scores (F(1,62) = 2.26, p = .044). As independent variable, VSA scores were significantly and positively associated with the perceived representativeness of the anatomy test questions (p = .010) and subjective improvement in anatomy knowledge of the lower extremity (p < .001).

## Sex differences

On baseline, males achieved significantly higher MRT scores than females (17.5 ± 4.9 vs 13.2 vs 4.9; p = .001). Both sexes performed equally well on written anatomical knowledge test (written knowledge test: 52.4 ± 18.9, p = .96; *factual:* 37.9 ± 20.4 vs 31.5 ± 17.9, p = .180; *functional:* 36.3 ± 17.4 vs 29.9 ± 12.7, p = .091; *spatial:* 51.4 ± 14.6 vs 47.1 ± 14.1, p = .242). However, males achieved significantly higher scores on the specimen test (51.5 ± 15.8 vs 40.0 ± 15.6; p = .005). Perceived cognitive load remained similar for both sexes (6.2 ± 1.2 vs 6.1 ± 1.1, p = .915).

# DISCUSSION

This study evaluated the effect of binocular disparity on learning anatomy in a stereoscopic 3D AR environment. Against author's expectations, no differences were found between stereoscopic 3D and monoscopic 3D view groups in terms of acquired anatomical knowledge and perceived cognitive load during learning. VSA, however, were significantly and positively associated with practical anatomical knowledge regardless of intervention. Additionally, VSA were positively associated with the perceived representativeness of anatomy test questions and the subjective improvement in anatomy knowledge of the lower extremity.

Although binocular disparity is generally considered to provide one of the important depth cues in 3D visualization, its exclusive effect on learning and cognitive load was revealed to be not significant in a stereoscopic 3D AR environment. To the author's knowledge only one study, performed by Wainman and colleagues, has evaluated the role of binocular disparity within the same type of technology.[40] Likewise, Wainman and colleagues found no beneficial effect of stereopsis on learning. The only difference between both studies was the way binocular disparity was eliminated. While in the current study a monoscopic view was obtained technically by presenting identical images to both eyes, Wainman and colleagues achieved monocular view by closing the dominant eye of participants with a patch. In addition, Wainman and colleagues have compared the effect of binocular disparity in AR to its effect in VR. The effect of stereoscopic vision in VR appeared to be significantly greater than in AR. In fact, learning with a stereoscopic 3D model in AR was less effective than in VR. This effect was explained by various degrees of stereopsis that different types of technologies can generate.

On the other hand, the findings suggest that other important depth cues could have compensated for the absence of stereopsis. During the experiment participants were able to walk around the 3D anatomical model and explore the model from all possible angles which is unique for a stereoscopic AR environment. This type of dynamic exploration, also referred to as motion parallax, is able to provide strong depth information.[48] Additional literature searches in the field of neurosciences education revealed that motion parallax in some cases can be even more effective than binocular disparity alone.[49-51] More interestingly, an interaction between both depth cues can exist.[52] For instance, subjects were asked to perform series of explorative tasks under three depth cue conditions: binocular disparity, motion parallax and combination of both depth cues.[50] The combination of binocular disparity and motion parallax resulted in an equal amount of correct answers as did the motion parallax condition (84% vs 80%; $p = .231$). However, in the absence of motion parallax, binocular disparity condition contributed to significantly less correct answers (60% vs. 80%; p < .001). In another study, that motion parallax improved performance in recovering 3D shape of objects in a monoscopic view, but not in a stereoscopic view.[53] Therefore, motion parallax could have reasonably compensated for the absence of binocular disparity and generated a sufficient 3D perception of the monoscopically projected 3D model. Further research is needed to evaluate to what extent motion parallax, alone and in combination with binocular disparity, affects learning.

Another effect of dynamic exploration, that could have occurred during this experiment, is the embodied cognition on learning.[54-56] Previous research has shown that using gestures and body movements helps students acquire anatomical knowledge. For instance, students who have engaged in miming using representational and metaphorical gestures while learning functions of central nervous system, have improved their marks with 42% in comparison with didactic learning.[56] Similar concept applies for mimicking specific joint movements in order to memorize them and being able to recall the structures names and to localize them on a visual representation.[55] Students in the current experiment were also using gestures while dissecting the anatomical layers and structures. That could have helped them memorizing structures while using similar gestures again and again. Additionally, students tended to move their own leg in a synchronized manner with the animated 3D model. Such engagement could have resulted in embodied learning and contribute to better learning within both modalities.

### The effect of VSA

In the current study anatomical knowledge was tested both by written and practical examinations. Both assessment methods were chosen to ensure a better alignment between learning and assessment of spatial knowledge. Consistent with previous research, VSA were positively associated with anatomical knowledge as measured by the practical specimen test.[45,57] However, VSA did not modify the observed outcomes as expected. In other words, individuals with lower VSA did not show different trajectory of learning with either monoscopic or stereoscopic 3D views of the model. Also, they did not experience significant differences in perceived cognitive load. This is in contrast with previous body of evidence on an aptitude-treatment effect caused by VSA when learning with different types of 3DVT.[23,28,58] If motion parallax was reasonably compensating for the absence of binocular disparity, as discussed above, then it does explain why students with lower VSA performed equally well within both conditions. These individuals were still able to generate proper 3D mental representations of the model within the monoscopic 3D view group and experienced equal amount of cognitive load during learning. Although the modifying effect of VSA on objective outcomes was not observed in current study, it was affecting the subjective outcomes regarding learning experience. This is particularly interesting, since the monoscopic 3D group with low VSA found the practical assessment items to be less representative of their learning environment than the stereoscopic 3D group.

Another explanation for the absence of modifying effect of VSA lies within the scale of spatial abilities needed for the task at hand. For spatial abilities a division between small- and large-scale space can be made, with small scale referring to space within arm's length, e.g., tabletop tasks. Large scale space refers to when locomotion is needed to interact with the spatial environment. As participants were walking around the model, large scale spatial processing takes place. As previously shown, a partial dissociation is found for small- and large-scale spatial abilities.[59] It could therefore be that the small-scale task of mental rotation used here, may not substantially relate to the large-scale spatial task of interacting with the model. Alternatively, large scale spatial tests, especially those relying on perspective taking, could show the interaction with task performance as hypothesized here.

Lastly, the observed sex differences in visual-spatial abilities scores in favor of males are in line with previous research.[60-64] More interestingly, males significantly outperform females on the specimen test, but not on the written knowledge test. Again, these findings suggest that the practical examination questions rely on visual-spatial abilities skills more than written knowledge test questions do. This is further supported by the work of Langlois and colleagues who have reviewed relationship between visual-spatial abilities test and anatomy knowledge assessment.[45] Authors have found significant relationship between spatial abilities test and anatomy knowledge assessment using practical examination, while relationship between spatial abilities an spatial multiple-choice questions remained unclear. Therefore, both findings suggest that practical examination questions are more reliable in testing spatial anatomical knowledge than multiple-choice questions, even when designed properly. Further research is needed to explore how spatial multiple-choice questions are mentally processed during examination in comparison to practical examination questions.

### Limitations

This study has several strengths and limitations. To authors' knowledge, this was the first single blinded randomized controlled trial to evaluate the effect of binocular disparity on learning in 3D AR environment within two academic centers and within one single level of instructional design. Along with the validated measurement instruments, it has maximized the internal and external validity of the results. On the other hand, participation was voluntary, and a selection bias could occur. The results could have been different if measured within the entire students' population. However, the baseline VSA scores among current study sample bear strong resemblance to VSA scores of the entire cohort of first-year medical undergraduates (14.9 vs. 14.4), as measured previously by Vorstenbosch and colleagues.[65] Another limitation was the relatively small sample size. Due to the limited availability of devices, authors were restricted to a maximum number of participants. It is possible that a much larger sample size could have revealed significant differences between interventions. The possible compensating effect of motion parallax and the effect of large-scale spatial abilities can also be considered as potential confounders that have influenced the internal validity. These new insights can help reveal the exact effect of both factors on learning. It is also important to note that the authors choose to not assess baseline knowledge to avoid extra burden for students and possible influence on learning during the intervention and performance on the post-tests. In this way any differences in prior knowledge that could have been present among students were not taken into account. Lastly, spatial knowledge questions in this study were carefully designed to stimulate mental visualizations skills. However, these questions can still be processed without spatial reasoning or just being best guessed when questions get too difficult to answer. Consequently, stereoscopic visualization of anatomy would not be that helpful in processing these type of questions.

### Future implications

The findings of this study have implications for both research and education. As stated previously, the aptitude treatment interaction caused by VSA should be taken into account when designing new research, especially when evaluating 3D technologies and their effect on learning. Additionally, the results of this study suggest that stereoscopic visualization can be differently effective depending on the type of technology used. More importantly, the findings suggest that other possible mechanisms are responsible for the acquired 3D effect and positive effect on learning. Next research should focus on the working mechanisms that explain the effectiveness of stereoscopic 3DVT. Only by knowing why particular 3D technology works will enable educators and researcher to properly design and implement this tool in medical education.

## CONCLUSIONS

In summary, binocular disparity alone does not contribute to better learning of anatomy in a stereoscopic 3D AR environment. Motion parallax, enabled by dynamic exploration, should be considered as a potential strong depth cue without or in combination with binocular disparity. Regardless of intervention, visual-spatial abilities were significantly and positively associated with the specimen test scores.

# REFERENCES

1. 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.
2. Peterson DC, Mlynarczyk GS. Analysis of traditional versus three-dimensional augmented curriculum on anatomical learning outcome measures. Anat Sci Educ 2016;9:529–536.
3. 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.
4. 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.
5. Bergman EM, de Bruin AB, Herrler A, Verhrijen IW, Scherpbier, van der Vleuten CP. Students' perceptions of anatomy across the undergraduate problem-based learning medical curriculum: A phenomenographical study. BMC Med Educ 2013;13:152–162.
6. 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.
7. Spielmann PM, Oliver CW. The carpal bones: A basic test of medical students and junior doctors' knowledge of anatomy. Surgeon 2005;3:257–259.
8. 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.
9. Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. How much anatomy is enough? Anat Sci Educ 2008;1:184–188.
10. Cook DA. The research we still are not doing: An agenda for the study of computer-based learning. Acad Med 2005;80:541–548.
11. Cutting JE, Vishton PM. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: Epstein W, Rogers S (Editors). Perception of Space and Motion. 1st Ed. San Diego, CA: Academic Press. 1995. p 69–117.
12. Johnston EB, Cumming BG, Landy MS. Integration of stereopsis and motion shape cues. Vision Res 1994;34:2259–2275.
13. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017;10:549–559.
14. Kytö M, Mäkinen A, Tossavainen T, Oittinen PT.Stereoscopic depth perception in video see-through augmented reality within action space. J Electron Imag 2014;23:011006.
15. Martinez EM, Junke B, Holub J, Hisley K, Eliot D, Winer E. Evaluation of monoscopic and stereoscopic displays for visual-spatial tasks in medical contexts. Comput Biol Med 2015;61:138–143.
16. Railo H, Saastamoinen J, Kylmala S, Peltola A. Binocular disparity can augment the capacity of vision without affecting subjective experience of depth. Sci Rep 2018;8:15798.
17. Anderson SJ, Jamniczky HA, Krigolson OE, Coderre SP, Hecker KG. Quantifying two-dimensional and three-dimensional stereoscopic learning in anatomy using electroencephalography. NPJ Sci Learn 2019;4:10.
18. Garg A, Norman GR, Spero L, Maheshwari P. Do virtual computer models hinder anatomy learning? Acad Med 1999;74:S87–S89.
19. Garg A, Norman G, Spero L, Taylor I. Learning anatomy: Do new computer models improve spatial understanding? Med Teach 1999;21:519–522.

20. 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.
21. Levinson AJ, Weaver B, Garside S, McGinn H, Norman GR. Virtual reality and brain anatomy: A randomised trial of e-learning instructional designs. Med Educ 2007;41:495–501.
22. Naaz F. Learning from graphically integrated 2D and 3D representations improves retention of neuroanatomy. University of Louisville: Louisville, KY. Doctorate of Philosophy Dissertation. 2002. 76 p.
23. Bogomolova K, van der Ham IJM, Dankbaar MEW, van den Broek WW, Hovius SER, van der Hage JA, Hierck BP. The effect of stereoscopic augmented reality visualization on learning anatomy and the modifying effect of visual-spatial abilities: a double-center randomized controlled trial. Anat Sci Educ 2020;13:558–567.
24. Hegarty M, Sims VK. Individual differences in use of diagrams as external memory in mechanical reasoning. Mem Cognit 1994;22:411–430.
25. Mayer RE. Multimedia Learning. 3rd Ed. Cambridge, UK: Cambridge University Press. 2001. 452 p.
26. Huk T. Who benefits from learning with 3D models? The case of spatial ability. J Comput Assist Learn 2006;22:392–404.
27. 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;6:211–215.
28. 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.
29. Kang X, Azizian M, Wilson E, Wu K, Martin AD, Kane TD, Peters CA, Cleary K, Shekhar R. Stereoscopic augmented reality for laparoscopic surgery. Surg Endosc 2014;28:2227–2235.
30. Cutolo F, Freschi C, Mascioli S, Parchi PD, Ferrari M, Ferrari V. Robust and accurate algorithm for wearable stereoscopic augmented reality with three indistinguishable markers. Electronics 2016;5:59.
31. Cutolo F, Meola A, Carbone M, Sinceri S, Cagnazzo F, Denaro E, Esposito N, Ferrari M, Ferrari V. A new head-mounted display-based augmented reality system in neurosurgical oncology: A study on phantom. Comput Assist Surg 2017;22:39–53.
32. Sommer B, Barnes DG, Boyd S, Chandler T, Cordeil M, Czauderna T, Klapperstück M, Klein K, Nguyen TD, Nim H, Stephens K. 3D-stereoscopic immersive analytics projects at Monash University and University of Konstanz. Electron Imag 2017:179–187.
33. Birt J, Stromberga Z, Cowling M, Moro C. Mobile mixed reality for experiential learning and simulation in medical and health sciences education. Information 2018;9:31.
34. Checcucci E, Amparore D, Pecoraro A, Peretti D, Aimar R, De Cillis S, Piramide F, Volpi G, Piazzolla P, Manfrin D, Manfredi M, Fiori C, Porpiglia F. 3D mixed reality holograms for preoperative surgical planning of nephron-sparing surgery: evaluation of surgeons' perception. Minerva Urol Nefrol. 2021;73:367-375.
35. Kumar RP, Pelanis E, Bugge R, Bruna H, Palomara R, Aghayan DL, Fretland AA, Edwin B, Elle OJ. Use of mixed reality for surgery planning: Assessment and development workflow. J Biomed Inform 2017;8:1000777
36. Schwab K, Smith R, Brown V, Whyte M, Jourdan I. Evolution of stereoscopic imaging in surgery and recent advances. World journal of gastrointestinal endoscopy 2017;9:368–377.

37. Itatani Y, Obama K, Nishigori T, Ganeko R, Tsunoda S, Hosogi H, Hisamori S, Hashimoto K, Sakai Y. Three-dimensional Stereoscopic Visualization Shortens Operative Time in Laparoscopic Gastrectomy for Gastric Cancer. Sci Rep 2019;9: 4108
38. Bogomolova K, Hierck BP, Looijen AEM, Pilon JNM, Putter H, Wainman B, Hovius SER, van der Hage JA. Stereoscopic three-dimensional visualisation technology in anatomy learning: A meta-analysis. Med Educ 2021;55(3):317-327.
39. Moro C, Phelps C, Redmond P, Stromberga Z. HoloLens and mobile augmented reality in medical and health science education: A randomised controlled trial. Br J Educ Technol 2021;52:680-694.
40. Wainman B, Pukas G, Wolak L, Mohanraj S, Lamb J, Norman GR. The critical role of stereopsis in virtual and mixed reality learning environments. Anat Sci Educ 2020;13:398–405
41. Cook DA, Beckman TJ. Reflections on experimental research in medical education. Adv Health Sci Educ Theory Pract 2010;15:455–464.
42. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills 1978;47:599–604.
43. 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.
44. Guillot A, Champely S, Batier C, Thiriet P, Collet C. Relationship between spatial abilities, mental rotation and functional anatomy learning. Adv Health Sci Educ Theory Pract 2007;12:491–507.
45. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ 2017;10:235–241.
46. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. 1956. 1st Ed. New York, NY: David McKay Company. 207 p.
47. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Adv Psychol 1988;52:139–83.
48. Rogers B, Graham M. Motion parallax as an independent cue for depth-perception. Perception 1979;8:125–134.
49. Bradshaw MF, Rogers BJ. The interaction of binocular disparity and motion parallax in the computation of depth. Vision Res 1996;36:3457–3468.
50. Naepflin U, Menozzi M. Can movement parallax compensate lacking stereopsis in spatial explorative search tasks? Displays 2001;22:157–164.
51. Aygar E, Ware C, Rogers D. The contribution of stereoscopic and motion depth cues to the perception of structures in 3D point clouds. ACM Trans Appl Percept 2018;15:9.
52. Lankheet MJ, Palmen M. Stereoscopic segregation of transparent surfaces and the effect of motion contrast. Vision Res 1998;38:659–668.
53. Sherman A, Papathomas TV, Jain A, Keane BP. The role of stereopsis, motion parallax, perspective and angle polarity in perceiving 3-D shape. Seeing Perceiving 2012;25:263–285.
54. Oh CS, Won HS, Kim KJ, Jang DS. “Digit anatomy”: A new technique for learning anatomy using motor memory. Anat Sci Educ 2011;4:132–141.
55. Cherdieu M, Palombi O, Gerber S, Troccaz J and Rochet-Capellan A. Make Gestures to Learn: Reproducing Gestures Improves the Learning of Anatomical Knowledge More than Just Seeing Gestures. Psychol 2017:1689.
56. Dickson AD and Stephens BW. It’s all in the mime: actions speak louder than words when teaching the cranial nerves. Anat Sci Educ 2015;8:584-592.

57. Roach VA, Mi M, Mussell J, Van Nuland SE, Lufler RS, DeVeau KM, Dunham SM, Husmann P, Herriott HL, Edwards DN, Doubleday AF, Wilson BM, Wilson AB. Correlating spatial ability with anatomy assessment performance: A meta-analysis. Anat Sci Educ 2021;14:317-329.
58. Luursema JM, Verwey WB, Kommers PA, Geelkerken RH, Vos HJ. Optimizing conditions for computer-assisted anatomical learning. Interact Comput 2006;18:1123–1138.
59. Hegarty M, Burte H, Boone AP. Individual differences in large-scale spatial abilities and strategies In: Montello DR (Editor). Handbook of Behavioral and Cognitive Geography. 2018. 1st Ed. Cheltenham, UK: Edward Elgar Publishing Limited; p 231–246.
60. Baenninger M, Newcombe N. The role of experience in spatial test performance: A meta-analysis. Sex Roles 1989;20:327–344.
61. 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.
62. 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.
63. 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.
64. Guimarães B, Firmino-Machado J, Tsisar S, Viana B, Pinto-Sousa M, Vieira-Marques P, Cruz-Correia R, Ferreira MA. 2019. The role of anatomy computer-assisted learning on spatial abilities of medical students. Anat Sci Educ 12:138–153.
65. Vorstenbosch MA, Klaassen TP, Donders AR, Kooloos JG, Bolhuis SM, Laan RF. Learning anatomy enhances spatial ability. Anat Sci Educ 2013;6:257–262..

# 5

# Stereoscopic video visualization of spatially-complex procedures increases performance of surgical novices with high visual-spatial abilities:

## a randomized controlled trial

Katerina Bogomolova, Jeroen J.G. van Merriënboer, Jan E. Sluimers,
Jeroen Donkers, Theo Wiggers, Steven E.R. Hovius, Jos A. van der Hage

*The American Journal of Surgery 2021; 222:739-745*

# ABSTRACT

**Background**

The effect of three-dimensional (3D) versus two-dimensional (2D) video on performance of a spatially complex procedure and perceived cognitive load were examined among residents in relation to their visual-spatial abilities (VSA).

**Methods**

In a randomized controlled trial, 108 surgical residents performed a 5-Flap Z-plasty on a simulation model after watching the instructional video either in a 3D or 2D mode. Outcomes included perceived cognitive load measured by NASA-TLX questionnaire, task performance assessed using Observational Clinical Human Reliability Analysis and the percentage of achieved safe lengthening of the scar.

**Results**

No significant differences were found between groups. However, when accounted for VSA, safe lengthening was achieved significantly more often in the 3D group and only among individuals with high VSA (OR=6.67, 95%CI: 1.23–35.9, p=.027).

**Conclusions**

Overall, 3D instructional videos are as effective as 2D videos. However, they can be effectively used to enhance learning in high VSA residents.

# INTRODUCTION

Surgical residents experience difficulties with learning and performing spatially-complex procedures that require spatial and conceptual understanding.[1] Consequently, residents tend to feel less confident about performing procedures with increasing complexity. In particular, individuals with lower visual-spatial abilities (VSA) experience difficulties in learning spatially complex procedures.[2,3] VSA is defined as the ability that allows individuals to construct visual-spatial, i.e., three-dimensional (3D), mental representations of two-dimensional (2D) images and to mentally manipulate these representations.[4,5] VSA has been found to be positively associated with both subjective and objective assessments of surgical performance, especially among novices.[6,7] In anatomical education, VSA has been widely explored and showed repeatedly its positive association with anatomical knowledge.[8]

The differences in performance between low and high VSA individuals are explained within the cognitive load theory (CLT)[9]. According to CLT, the capacity of human working memory involved in processing new information is severely limited. Types of cognitive load include intrinsic (caused by performing the task itself), extraneous (caused by the way learning material is presented), and germane (caused by actual learning) load[9]. When the sum of the three types of load exceeds working memory capacity, a cognitive overload occurs which impairs learning.[10] Since low VSA individuals devote more cognitive resources to performing a spatially complex task, their intrinsic cognitive load increases during learning. Subsequently, they are left with less available resources that they can allocate to learning which leads to decreased performance in comparison with high VSA individuals. One way to compensate for the increased cognitive load in low VSA individuals, is to decrease extraneous load by improving the instructional method.

Instructional videos are one of the most used and effective ways of preparing for surgeries among residents.[11,12] When presented in a segmented rather than continuous format, using a step-by-step approach, video-based learning can be even more effective.[13] However, procedures are viewed monoscopically on a 2D screen without real perception of visual depth. This can make learning spatially-complex procedures more challenging, especially for residents with lower levels of VSA, as they have less ability to transform 2D images into 3D mental representations. However, according to the compensating hypothesis within the CLT,[10] this mental transformation can be assisted by presenting images stereoscopically, in real 3D.[14,15] In other words, stereoscopic visualization can compensate for low VSA by providing depth cues. In this case, the mental 3D model is already built and provided. This can eventually lead to a decreased cognitive load and improved learning.[15,16] A binocular vision of the viewer, though, is required to perceive spatial visual depth that is obtained with the use of this technology.

Compared to a monoscopic visualization, stereoscopic visualization has been shown to improve performance in laparoscopic surgery training.[17,18] However, it is yet unknown whether this applies to video-based learning and would probably depend on the “spatial” nature of the procedure. Roach et al. evaluated the effectiveness of a stereoscopic 3D instructional video of a four-flapped Z-plasty and a rhomboid flap and found no significant differences in performance when compared to a monoscopic 2D video of the same procedures.[19] As stated by the authors, both procedures were “essentially two dimensional in their design and performance, leaving their complexity to appear more conceptually geared rather than spatially.” This suggests that watching an instructional video in real 3D that is eventually performed in a 2D plane does not offer extra gain in terms of knowledge and performance skill. This also suggests that video demonstration of a spatially complex procedure performed in a 3D plane can benefit from stereoscopic visualization, given that the complexity of the procedure is based more on spatial thinking and understanding.

Therefore, the primary aim of this study was to evaluate whether a 3D instructional video compared to a 2D video of a 5-flap Z-plasty would improve performance of surgical residents. The secondary aim was to evaluate the perceived cognitive load both during watching the instructional video and performing the procedure. The outcomes were evaluated in relation to the VSA of participants.

# METHODS

### Study design and population

A randomized controlled trial was conducted at the Leiden University Medical Center, The Netherlands (Figure 1). The experiment took place during a hands-on session as part of a special educational surgery program in September 2019. The study protocol was approved by the Netherlands Association for Medical Education (NVMO) Ethical Review Board (NERB case number: 2019.6.10).

### Participants

Participants were first-year surgical residents from various teaching hospitals in The Netherlands. Participation in this study was voluntary, and written consent was obtained from all enrolled residents.

### Randomization

Participants were randomly allocated to either a 2D or 3D video group using an Excel Random Group Generator (Microsoft Excel for Office 365 MSO, version 2012). Randomization was stratified by sex to ensure an equal female-to-male ratio in both groups.

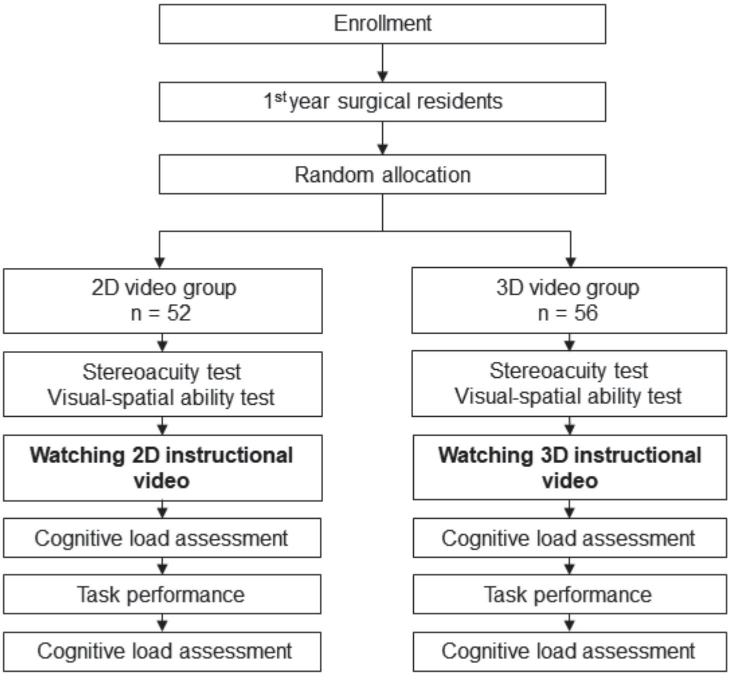


Figure 1. Flowchart of study design. 2D, two-dimensional; 3D, three-dimensional; n, number of participants.

### Stereovision

Stereovision of participants was measured by a Random Dot 3 - LEA SYMBOLS® Stereoacuity Test [Vision Assessment Corp., Elk Grove Village, IL, USA] prior to the experiment to identify and exclude individuals with absent stereovision.

### Assessment of VSA

VSA was measured by a validated Mental Rotation Test (MRT).[20,21] MRT is the gold standard to assess VSA that has been associated with anatomical knowledge and surgical skills assessment.[7,22] The duration of the test was 10 min, and the test scores ranged from 0 to 24 points. High VSA was defined as a mean score above the average; and low VSA was defined as a mean score below the average.

### Surgical procedure

A 5-flap Z-plasty, also referred to as the jumping man flap, is a spatially complex procedure that involves two different types of tissue movements in a 3D plane. Its goal is to simultaneously provide lengthening and deepening of a skin contracture. The lengthening is achieved by two Z-plasties, while the deepening is achieved by a V-Y advancement flap (*Supplementary material 8).* Spatial understanding of the problem and solution is required to perform the procedure correctly.

### Instructional videos

The instructional video of a 5-flap Z-plasty was developed using a validated step-by-step approach.[12] The video consisted of a step-by-step demonstration of the procedure on a simulation model accompanied by auditory narration. The duration of the video was 4 min.

The 2D video group watched the instructional video on a large flat screen. The 3D video group watched the same video on a large flat screen with stereoscopic projection and the use of active 3D glasses. All participants watched the video twice to stimulate active processing. Participants were asked to mentally rehearse the surgical steps while watching the video the second time without auditory narration.

### Cognitive load assessment

Perceived cognitive load was measured twice by a validated NASA-TLX questionnaire.[23] The participants filled out the questionnaire immediately after watching the instructional video and for the second time, after performing the task. The cognitive domains included mental, physical, and temporal demands; performance; effort; and frustration. The total score was a product of the six domains and ranged between 0 and 10 points.

### Task performance

Participants performed a 5-flap Z-plasty on a skin contracture in the simulation model (Figure 2) after watching the instructional video. The task was to execute the most favorable advanced 5-flap Z-plasty given the initial length of contracture and maximal slack of the skin to achieve maximal lengthening and deepening, taking into account the viability of the skin flaps. Participants were given 15 min to complete the procedure.

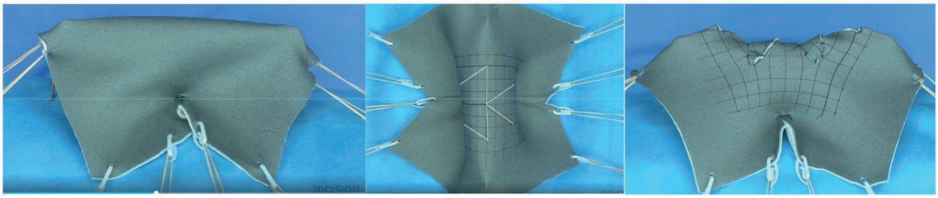


Figure 2. Simulation model of a skin contracture. Left: front view of the skin model, middle: above view of the skin with a 5-flap Z-plasty being drawn, right: the final result.

### OCHRA checklist

The performed procedures were assessed independently by two experts (JS and KB) in a blinded manner using the OCHRA checklist (*Supplementary material 9*). The OCHRA is a procedure-specific, step-by-step skills assessment checklist that is characterized by a breakdown of a procedure into tasks or substeps.[24] The 5-flap Z-plasty comprises 10 substeps, each of which can attain scores of 0.5, 1, or 1.5 points if performed correctly. The total maximum possible score was 10 points.

### Safe lengthening

The safe range of minimal and maximal gain in lengthening that can be achieved, if performed correctly, was calculated for each performed procedure given the initial length of the contracture and the chosen size of the angles of the Z-plasty flaps. The calculations were performed blindly by two experts (KB and JS) after the experiment. Safe lengthening was achieved (yes/no) if the calculated value felt in the permitted range of lengthening without compromising blood flow of the created flaps.

### Outcome measures

The primary outcome measure was defined as the difference in mean OCHRA scores between 2D and 3D video groups. Secondary outcomes were defined as the difference in proportions of achieved safe lengthening and mean cognitive load scores of instructional videos, and task performance between the two groups. Additionally, the outcomes were evaluated for different levels of VSA.

### Statistical analysis

Owing to the novelty of the study, including the type of procedure and assessment tool, no previous data were available to calculate the sample size. A minimal sample size of 100 participants was assumed to be appropriate. Participants' baseline characteristics were summarized using descriptive statistics. The differences in baseline measurements were assessed using an independent *t*-test for differences in means and chi-squared test for differences in proportions. The differences in mean OCHRA and cognitive load scores were assessed with an independent *t*-test, and the proportions of safe lengthening, with a chi-squared test. To assess the possible modifying effect of VSA on outcomes, regression analyses were performed. In ANCOVA, the OCHRA and cognitive load scores were included as dependent variables, intervention group as the fixed factor, VSA score as the covariate, and 'intervention x VSA' as the interaction term. In the logistic regression model, safe lengthening was included as the dependent variable; the remaining factors were identical as for the ANCOVA test. Odd ratios were calculated based on the values of predictors and interaction term from the logistic regression model. 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

A total of 108 participants were included (Table 1). All participants could perceive spatial-visual depth as measured by the stereoacuity test.

The differences between the 2D and 3D video groups are presented in Table 2. The performance on the task, as measured by the OCHRA checklist, was not significantly different between the two groups (t(106) = .813; p = .487). Although safe lengthening was achieved more often by the participants in the 3D group, these differences were not significant (56.8% vs. 43.2%; $c^2(1) = .602$; p = .555). The perceived cognitive load was similar in both groups.

### The effect of VSA

The mean OCHRA scores remained similar for all levels of VSA, as measured by the interaction term in ANCOVA (F(1, 102)=0.43, p = .513). However, regardless of intervention, VSA was significantly and positively associated with performance only among individuals with initially low levels of VSA (F(1,48) = 5.37, partial $\eta^2 = 0.11$, p = .025) (Figure 3a). This association was not found among individuals with high VSA (F(1, 43) = 0.26, partial $\eta^2 = 0.006$, p = .610) (Figure 3b).

Table 1. Baseline characteristics of the included participants.

| | 2D video | 3D video | p value |
|---|---|---|---|
| | n = 52 | n = 56 | |
| Sex, n (%) | | | |
| Male | 25 (48.1) | 26 (46.4) | .622 |
| Female | 27 (51.9) | 30 (53.6) | |
| Age, mean ±SD in years | 29.8 ± 2.09 | 30.5 ± 1.5 | .053 |
| Residency, n (% | | | |
| General | 26 (50.0) | 34 (60.7) | .571 |
| Orthopedic | 13 (25.0) | 10 (17.9) | |
| Plastic and Reconstructive Surgery | 4 (7.7) | 6 (10.7) | |
| Urology | 9 (17.3) | 6 (10.7) | |
| Clinical experience, n (%) | | | |
| Yes | 50 (96.2) | 56 (100) | .210 |
| No | 2 (3.8) | 0 (0) | |
| Clinical experience, mean ± SD in months | 21.3 ± 11.3 | 20.8 ± 8.8 | .801 |
| OR Hours, median (IQR) | 1.5 (4.8) | 1.0 (6.4) | .588 |
| Performed a Z or V-Y plasty before, n (%) | | | |
| Yes | 6 (11.5) | 4 (7.1) | .414 |
| No | 46 (88.5) | 52 (92.9) | |
| Mental Rotation Test score, mean ± SD | 13.2 ± 5.4 | 13.8 ± 5.5 | .596 |

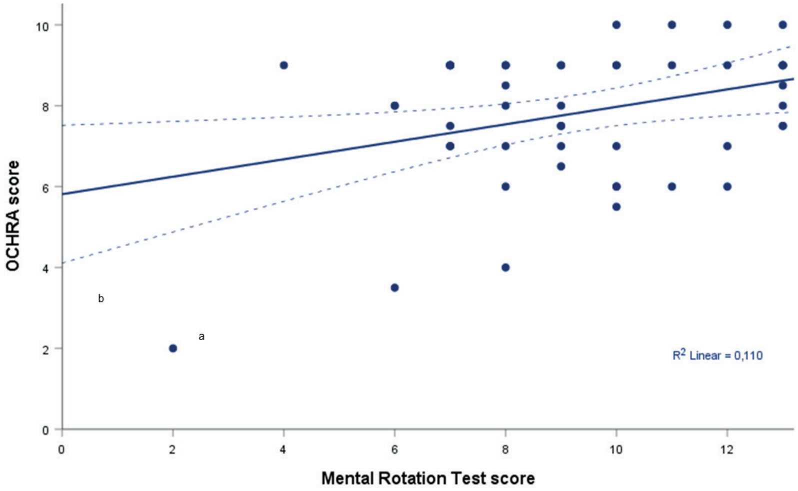
2D, two-dimensional; 3D, three-dimensional; n, number of participants; SD, standard deviation; OR, operating room; IQR, interquartile range

Table 2. Differences in outcomes between 2D and 3D instructional video groups.

| | 2D video | 3D video | p value |
|---|---|---|---|
| | n = 52 | n = 56 | |
| OCHRA score, mean ± SD | 8.0 (1.7) | 7.8 (2.0) | .487 |
| Safe lengthening, n (%) | | | |
| Achieved | 19 (43.2) | 25 (56.8) | .555 |
| Not achieved | 32 (50.8) | 31 (49.2) | |
| Cognitive load video, mean ± SD | 6.1 ± 1.2 | 6.0 ± 1.4 | .738 |
| Cognitive load task, mean ± SD | 6.5 ± 1.0 | 6.1 ± 1.2 | .142 |

2D, two-dimensional; 3D, three-dimensional; n, number of participants; SD, standard deviation.

a. Low VSA group



b. High VSA group

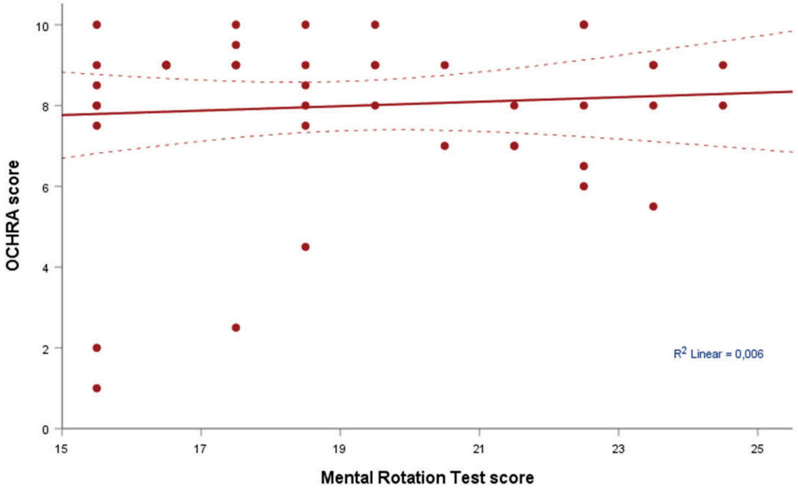


Figure 3. The relationship between Mental Rotation Test (MRT) scores and OCHRA scores in the low VSA (a) and high VSA (b) groups. The continuous lines represent the prediction values of the OCHRA scores; the dashed lines represent the 95% confidence Interval of these values.

The proportions of achieved safe lengthening were significantly different for participants with low and high VSA, as measured by the interaction term in logistic regression (OR = 6.67, 95% CI: 1.23–35.9, p = .027). Safe lengthening was achieved significantly more often in the 3D group and only among individuals with high VSA than in the 2D group (OR = 8.8). The corresponding odds ratios (OR) are illustrated in Figure 4. The coefficients' estimates for the model are provided in *Supplementary material 10*.

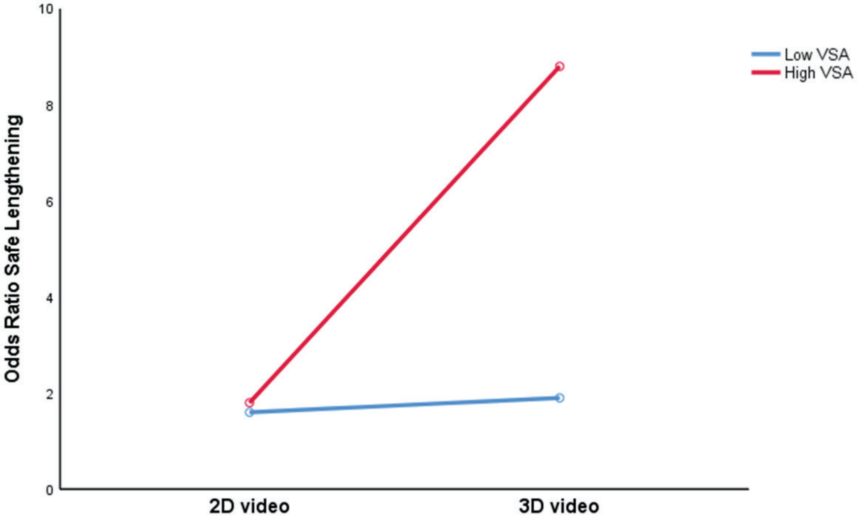


Figure 4. The interaction between intervention (2D versus 3D video group) and visual-spatial abilities (low versus high VSA) for the outcome "safe lengthening." The OR was calculated based on the values of predictors and interaction term in the logistic regression model.

The perceived cognitive load scores remained similar for all levels of VSA in both study groups, as measured by the interaction term in ANCOVA (*video:* F(1, 100) = 1.81, p = .182; *task:* F(1,97) = 0.78, p = .379). Regardless of intervention, VSA was not associated with the perceived cognitive load.

# DISCUSSION

In this study, the effectiveness of a 3D versus 2D instructional video of a spatially complex procedure was evaluated in terms of performance and perceived cognitive load among surgical residents. Additionally, the outcomes were evaluated in relation to VSA. The beneficial effect of the stereoscopic video visualization hypothesized to be the greatest for individuals with lower VSA was not observed. On the contrary, the learning effect of the 3D mode was greater among individuals with higher VSA than in those with lower VSA.

These findings are in contrast to the proposed compensating hypothesis assuming that presenting images stereoscopically would compensate for low VSA, as the mental 3D model is already built and provided.[15,16] This eventually would lead to a decreased cognitive load for low VSA learners and improve learning. However, the opposite effect was observed with high VSA learners benefiting most from 3D visualization. This interaction could have been caused by the so-called "expertise reversal effect" that occurs between different levels of learner's expertise and instructional techniques.[25,26] Novices with minimal prior knowledge must process many novel elements of information that can easily overload their working memory capacity. Accordingly, instructional methods that are effective for experienced learners may become less effective or even disadvantageous for novice learners, and vice versa. It has been reported that learners with high prior knowledge benefit more from continuous dynamic animations, while novices with no or low prior knowledge can have difficulties with processing high degrees of transitivity of visual presentations.[27,28] Because the participants of this study were novices in terms of most surgical procedures with an average of 1.5 hours spent in OR a week, the instructional video could have been perceived as very challenging, which was reflected by the relatively large proportion of participants (58.3%) that did not achieve safe lengthening and the relatively high perceived cognitive load. The direction of this interaction is in line with the co-existing mechanism, the ability-as-enhancer hypothesis, within the cognitive load theory.[29,30] This hypothesis predicts that high levels of VSA are required to benefit from improved instructional method, while low VSA levels are hindered by it. In other words, high levels VSA enhance learning with 3D visualization, while low VSA levels do not. This is because high VSA individuals have still sufficient number of cognitive resources after processing spatially complex procedure. Low VSA individuals, on the other hand, are depleted in their cognitive resources and are not able to benefit from learning with 3D visualization.

Notably, the interaction effect of VSA and performance, as measured by the OCHRA checklist, was not observed. This can be explained by the choice of the assessment tool. In research and practice, surgical performance is effectively assessed both by global rating scales and task-specific checklists.[31] A task-specific OCHRA checklist was preferred in the current study for several reasons. First, the OCHRA checklist allowed individual assessment of each performed step of the 5-flap Z-plasty. This was essential, because each step of the procedure requires spatial understanding to be performed correctly, and this can affect the final result. Second, global rating scales include items that are much less relevant in performing a 5-flap Z-plasty, such as knowledge and use of instruments, and respect for other tissues.[32]. However, the association between VSA and performance appears to be most prominent when performance is measured by a global rating scale. Wanzel et al. evaluated the effect of VSA, as measured by the MRT, on performance of a spatially complex procedure.[3] Performance was measured both by a global rating scale and task-specific checklist. The MRT scores were significantly associated with the scores on the global rating scale (r = 59, p = .0013), but not with the checklist scores (r = 0.36, p = .068). This can explain why the interaction effect between MRT and OCHRA scores was not observed in the current study.

Regardless of intervention, VSA was associated with performance only among individuals with lower VSA levels. The specific VSA scores of learners higher on the VSA continuum did not affect their surgical performance. A similar association was previously described by Roach et al. evaluating the effect of stereoscopic visualization on laparoscopic skills in relation to VSA.[33] This suggests that VSA may impair performance if not well developed, but VSA does not affect performance when a certain VSA level has been reached. Previous research has shown that VSA can be trained by repeated practice and learning.[34-36] Therefore, training of VSA skills in novices with initially low VSA could be beneficial.

### Limitations

This study has some limitations. First, all participants watched the instructional video simultaneously on a large computer screen. Watching it at their own pace was therefore not possible. It has been reported that pauses provide learners more time to process, consolidate, and transfer information to long-term memory before moving forward to the next step. Moreover, they can also replay the steps and better process the difficult parts of the procedure.[37] As our study participants were novices, those subjects with low VSA could have benefited from self-paced control and perhaps, even benefited from 3D visualization. Second, the included participants were first-year surgical residents with little to no prior surgical knowledge and experience. As more complex instructional videos can be more effective for more experienced learners, better effects of 3D visualization could have been obtained in the later stages of surgical training. Third, 5-flap Z-plasty was performed on a simulated model that optimally, but not fully, resembled normal skin. The used material could have permitted a larger amount of stretch of the flaps than normal skin. This could have led to an unintended increase in the achieved lengthening and deepening of the contracture. To account for the possible bias, the achieved safe lengthening was calculated for each participant. The assessment of achieved deepening was not possible owing to variability in height of the contracture. Last, all participants could complete the procedure within the given time frame of 15 min. Differences in time to complete the procedure between groups were not evaluated. However, by including time as an outcome measure, more insight could have been gained into the effects of VSA on performance.

### Future implications

The findings of this study have implications for both research and practice. When designing a new study, it is instrumental to consider VSA as a potential effect modifier. Additionally, a potential interaction between VSA and instructional design should be considered when performing statistical analysis. In surgical practice, an individualized approach could be helpful for residents with lower levels of VSA. Individuals with low VSA can benefit from deliberate practice and feedback and achieve a comparable level of competency as those with high VSA.[2] Because spatial skills are malleable, VSA-based training is recommended in the early stages of surgical training to reinforce the development of surgical skills among novices.[34-36] For individuals with low VSA, 3D instructional videos seem to initially have no advantage over conventional 2D videos. Further research is encouraged to evaluate whether instructional videos are more effective in the later stages of surgical training and whether low VSA learners will then benefit from 3D visualizations.

# CONCLUSIONS

Overall, watching an instructional video of a spatially complex procedure in 3D is as effective as watching it in a 2D traditional format. However, when considering VSA of the learners, 3D visualization is more favorable for individuals with high VSA. Authors hypothesized that the educational effect of 3D visualization possibly depends not only on VSA but also on the expertise level of the learner and/or the complexity of the procedure. Future research should focus on the effect of 3D instructional videos of simpler vs. complex procedures for residents with low VSA in relation to their level of expertise.

# REFERENCES

1. Klebanoff JS, Marfori CQ, Vargas MV, et al. Ob/Gyn resident self-perceived preparedness for minimally invasive surgery. BMC Med Educ. 2020;20(1):185.
2. Wanzel KR, Hamstra SJ, Anastakis DJ, et al. Effect of visual-spatial ability on learning of spatially-complex surgical skills. Lancet. 2002;359(9302):230-231.
3. Wanzel KR, Hamstra SJ, Caminiti MF, et al. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. Surgery. 2003;134(5):750-757.
4. Gordon HW. The cognitive laterality battery: tests of specialized cognitive function. Int J Neurosci. 1986;29(3-4):223-244.
5. Kozhevnikov M, Hegarty M. A dissociation between object manipulation spatial ability and spatial orientation ability. Mem Cognit. 2001;29(5):745-756.
6. Maan ZN, Maan IN, Darzi AW, Aggarwal R. Systematic review of predictors of surgical performance. Br J Surg 2012;99(12):1610-1621.
7. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and technical skills performance in health care: a systematic review. Med Educ. 2015;49(11):1065-1085.
8. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial Abilities Training in Anatomy Education: A Systematic Review. Anat Sci Educ. 2020;13(1):71-79.
9. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ. 2010;44(1):85-93.
10. Sweller J, van Merrienboer JJG, Paas F. Cognitive architecture and instructional design: 20 years later. Educ Psychol Rev. 2019;31(2):261-292.
11. Green JL, Suresh V, Bittar P, et al. The utilization of video technology in surgical education: A systematic review. J Surg Res. 2019;235:171-180.
12. Ahmet A, Gamze K, Rustem M, Sezen KA. Is video-based education an effective method in surgical education? A systematic review. J Surg Educ. 2018;75(5):1150-1158.
13. Nazari T, van de Graaf FW, Dankbaar MEW, et al. One step at a time: step by step versus continuous video-based learning to prepare medical students for performing surgical procedures. J Surg Educ. 2020;7(4):779-787
14. Hoffler TN. Spatial ability: Its influence on learning with visualizations-a meta-analytic review. Educ Psychol Rev. 2010;22(3):245-269.
15. Mayer RE. Multimedia Learning. Cambridge University Press: Cambridge, 2001.
16. Lee DY, Shin DH. An empirical evaluation of multi-media-based learning of a procedural task. Comput Hum Behav. 2012;28(3):1072-1081.
17. Fergo C, Burcharth J, Pommergaard HC, et al. Three-dimensional laparoscopy vs 2-dimensional laparoscopy with high-definition technology for abdominal surgery: a systematic review. Am J Surg. 2017;213(1):159-170.
18. Sorensen SM, Savran MM, Konge L, Bjerrum F. Three-dimensional versus two-dimensional vision in laparoscopy: a systematic review. Surg Endosc. 2016;30(1):11-23.
19. Roach VA, Brandt MG, Moore CC, Wilson TD. Is three-dimensional videography the cutting edge of surgical skill acquisition? Anat Sci Educ. 2012;5(3):138-145.
20. Shepard RN, Metzler J. Mental rotation of three-dimensional objects. Science. 1971;171(3972):701-703.
21. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills. 1978;47(2):599-604.

22. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ. 2017;10(3):235-241.
23. Hart S, Staveland L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Adv Psychol. 1988;52:139-183.
24. Nazari T, Vlieger EJ, Dankbaar MEW, et al. Creation of a universal language for surgical procedures using the step-by-step framework. BJS Open. 2018;2(3):151-157.
25. Chen OH, Kalyuga S, Sweller J. The expertise reversal effect is a variant of the more general element interactivity effect. Educ Psychol Rev. 2017;29(2):393-405.
26. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. Educ Psychol-US. 2003;38(1):23-31.
27. Kalyuga S. Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. Comput Hum Behav, 2008;24(3):852-861.
28. Kalyuga S, Rikers R, Paas F. Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. Educ Psychol Rev. 2012;24(2):313-337.
29. Huk T. Who benefits from learning with 3D models? The case of spatial ability. J Comp Assisted Learning. 2006;22(6):392-404.
30. Mayer RE, Sims VK. For whom is a picture worth 1000 words - extensions of a dual-coding theory of multimedia learning. J Educ Psychol. 1994;86(3):389-401.
31. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, et al. Objective assessment of technical surgical skills. Brit J Surg. 2010;97(7):972-987.
32. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273-278.
33. Roach VA, Mistry MR, Wilson TD. Spatial visualization ability and laparoscopic skills in novice learners: evaluating stereoscopic versus monoscopic visualizations. Anat Sci Educ. 2014;7(4):295-301.
34. Uttal DH, Meadow NG, Tipton E, et al. The malleability of spatial skills: a meta-analysis of training studies. Psychol Bull. 2013;139(2):352-402.
35. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities training in anatomy education: A systematic review. Anat Sci Educ. 2020;13(1):71-79.
36. Bogomolova K, Hierck BP, van der Hage JA, Hovius SER. Anatomy dissection course improves the initially lower levels of visual-spatial abilities of medical undergraduates. Anat Sci Educ. 2020; 13(3):333-342.
37. Schwan S, Riempp R. The cognitive benefits of interactive videos: learning to tie nautical knots. Learn Instr. 2004;14(3):293-305.

# 6

# Global versus task-specific post-operative feedback in surgical procedure learning:
## a randomized controlled trial

Tahmina Nazari* & Katerina Bogomolova*, Marlike Ridderbos,
Mary E.W. Dankbaar, Jeroen J.G. van Merriënboer, Johan F. Lange,
Theo Wiggers, Jos A. van der Hage
* Shared first authorship

*Surgery 2021; 170:81-87*

# ABSTRACT

**Background**

Task-specific checklists and global rating scales are both recommended assessment tools to provide constructive feedback on surgical performance. This study evaluated the most effective feedback tool by comparing the effects of the Observational Clinical Human Reliability Analysis (OCHRA) and the Objective Structured Assessment of Technical Skills (OSATS) on surgical performance in relation to the visual-spatial abilities (VSA) of the learners.

**Methods**

In a randomized controlled trial, medical students were allocated to either the OCHRA (n = 25) or OSATS (n = 25) feedback group. VSA was measured by a Mental Rotation Test (MRT). Participants performed an open inguinal hernia repair procedure on a simulation model twice. Feedback was provided after the first procedure. Improvement in performance was evaluated blindly using a global rating scale (performance score) and hand-motion analysis (time and path length).

**Results**

Mean improvement in performance score was not significantly different between the OCHRA and OSATS feedback groups (p = .100). However, mean improvement in time (371.0 ± 223.4 vs 274.6 ± 341.6; p = .027) and path length (53.5 ± 42.4 vs 34.7 ± 39.0; p = .046) was significantly greater in the OCHRA feedback group. When stratified by MRT scores, the greater improvement in time (p = .032) and path length (p = .053) was observed only among individuals with low VSA.

**Conclusions**

A task-specific (OCHRA) feedback is more effective in improving surgical skills in terms of time and path length in novices compared to a global rating scale (OSATS). The effects of a task-specific feedback are present mostly in individuals with lower VSA.

# INTRODUCTION

Feedback has long been recognized for its positive effect in surgical knowledge and skills training.[1] It has been shown to be crucial in technical skill development because it increases motivation, prevents incorrect actions, and reinforces correct actions.[2, 3] Feedback can be provided based on direct observation of technical skills.[4] Within the surgical field, different observational assessment tools are available.[5] Assessment tools assess surgical performance on competences, skills, or surgical-specific items on a checklist. These tools can be used as a medium for feedback to provide information regarding a trainee's performance to improve on specific items that are being assessed.[1,5] Two main types of assessment tools can be recognized: global rating scales, which rate general surgical skills and are applicable to all surgical procedures, or procedure-specific checklists.[5] In both categories, many tools have been developed and validated.[4,5]

A commonly used and generally accepted as "gold standard" assessment tool is Objective Structured Assessment of Technical Skills (OSATS), a global rating scale introduced by Martin et al for assessing technical skills of an entire surgical procedure.[5, 6] OSATS is a reliable, validated tool that assesses 7 competencies on a 5-point Likert scale.[6] It is feasible and effective in assessment of surgical skills of trainees in the operating room.[7]

Although global rating scales such as the OSATS are easy in use, these scales can be imprecise.[4] A task-specific method may provide more concise and precise feedback.[4] A task-specific technical skills assessment method is the Observational Clinical Human Reliability Analysis (OCHRA).[8] An OCHRA checklist assesses in a stepwise manner whether a substep was correct or incorrect.[8] Both OSATS and OCHRA assessment tools have shown to be valid for providing constructive feedback.[4,7] However, according to constructive alignment theory, the OCHRA feedback might be more effective when the surgical procedure is also learned in a stepwise manner.[9]

Although the validity of OSATS and OCHRA is demonstrated, these assessment tools are still based on individual judgments, which are inevitably associated with subjectivity.[10] Quantifying measures of technical skills may potentially mitigate this subjectivity. For open surgery, different motion tracking devices are described to measure either hand or instrument movements.[11-14] The outcomes of time to complete a task and total path length can differentiate between novices and experts.[13-15]

Additionally, the effect of feedback in relation to visual-spatial abilities (VSA), as another determining factor for technical skills development, is unrecognized. VSA are defined as the ability that allows individuals to construct visual-spatial (i.e., 3-dimensional) mental representations of 2D images and to mentally manipulate these representations.[16,17] This ability determines how well individuals are able to translate the acquired anatomical knowledge into clinical and surgical practice. Consequently, VSA determine how well surgical residents can understand and perform spatially complex procedures. The positive association between VSA and acquisition of surgical skills, including quality of hand motion, has been observed especially in the early phases of surgical training.[15,18-20] Moreover, VSA can have a modifying effect on outcomes. Individuals with lower VSA tend to perform worse than individuals with high VSA on acquisition of anatomical knowledge and surgical skills. However, with supportive instructional methods and deliberate practice and feedback they are able to achieve a comparable level of competency.[15,21- 23]

The aim of this study was to investigate whether a task-specific, stepwise feedback checklist (OCHRA) leads to a greater improvement in performance of a surgical procedure compared to a global rating scale method (OSATS) in terms of improvement of overall performance score, time to complete task, and total path length. These outcomes were also evaluated in relation to learners' VSA.

# METHODS

### Study design and population

A randomized controlled trial was conducted at the Leiden University Medical Center, The Netherlands. Participants were medical students and novices to almost any type of surgical procedures. Only right-handed students were included because left-handed novice students may have difficulties with the surgical instruments.[24] Participation was voluntary, and written consent was obtained from all participants. The study protocol was approved by the Netherlands Association for Medical Education Ethical Review Board (NERB dossier number: 1013) (Figure 1).

### Randomization

Participants were randomly allocated to either the OCHRA feedback (n = 25) or OSATS feedback group (n = 25) using an Excel random group generator.

### Surgical procedure

The Lichtenstein open inguinal hernia repair was chosen as a procedure containing multiple surgical steps and because of its spatial complexity, which requires a certain level of surgical anatomical knowledge and VSA of the learner. The first part of the surgery, until resecting the hernia sac, requires solely basic surgical skills such as incising, dissecting, and ligating. The second part, the placement and fixation of the mesh, is more complex. Each participant performed the Lichtenstein open inguinal hernia repair 2 times on a validated simulation model.[25] Participants were given access to the online course 1 week before the experiment to prepare for the experiment. The course consisted of 3 components: an introductory description that included text and figures regarding the surgical anatomy, a stepwise textual description, and a video demonstration of the procedure on the identical model used during the experiment.[26] The video demonstration depicted all important steps that need to be undertaken during surgery. Video was accompanied by auditory explanation. Participants were able to retrieve the materials as many times as they wanted and were able to do it on their own pace.

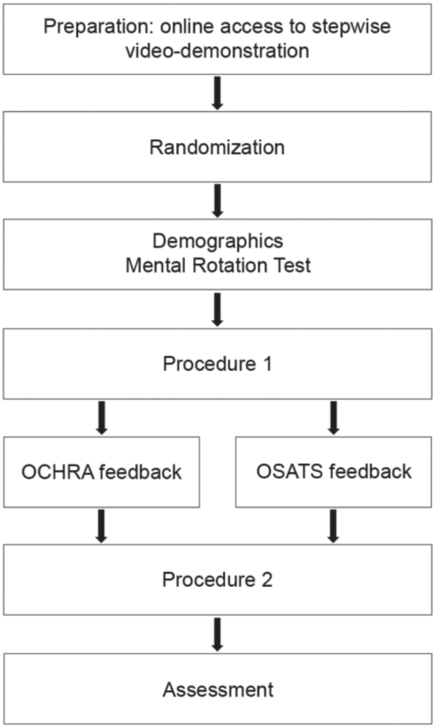


Figure 1. Flowchart of study design.

On the day of experiment, participants were given 30 minutes to complete each procedure.[27] The second procedure was performed directly after the provided feedback on the first procedure. Both procedures were recorded on video for blinded assessment. Participants were wearing a right-hand glove for the recording of motion by a motion tracking device (PST Base, PS-Tech B.V., Amsterdam, The Netherlands).

### Demographic questionnaire

The questionnaire was administered before the experiment to account for factors that could possibly influence the performance. In a previous study, the time students studied for the open inguinal hernia repair with the use of a video demonstration had a significant modifying effect on surgical performance.[27] Therefore, study time was included in the questionnaire and was accounted for in the data analysis.

### Assessment of VSA

VSA were measured by the mental rotation test (MRT) before the experiment. The MRT is a validated 24-item psychometric test and is the gold standard in assessing VSA in anatomical and surgical education.[19,28, 29, 30] Participants were given 10 minutes to complete the test. The maximum possible score for the test was 24 points.

### Interventions

In the OCHRA feedback group, postoperative feedback was provided using OCHRA. The OCHRA checklist is a reliable and valid instrument that has been successfully used in assessment of performance in various surgical procedures.[8,31, 32, 33] It is a procedure-specific step-by-step skills assessment checklist that is characterized by a breakdown of a procedure into tasks.[26] Each step is assessed for being performed correctly and if errors are being made during the particular step. Provided feedback was based on the evaluation of each performed procedural step (*Supplementary material 11).* If a particular step was performed incorrectly, the error was discussed, and a proper execution of the step was explained. No points or final scores were awarded for the performance.

In the OSATS feedback group, postoperative feedback was provided using the OSATS assessment tool (*Supplementary material 12)* OSATS is a 7-item global rating scale that focuses on the following overall competencies: (1) respect for tissue, (2) time and motion, (3) instrument handling, (4) knowledge of instruments, (5) use of assistance, (6) flow of operation, and (7) knowledge of procedure.[6] The tool has been previously validated in a wide range of surgical procedures and disciplines with reasonable index of reliability.[6,34,35] Provided feedback was based on the evaluation of each of the 7 competencies in the exact order of OSATS. Suboptimal performance and errors made within a competence were discussed based on an example followed by an explanation for the improvement. No points or final scores were awarded for the performance to avoid any bias that could be introduced by grading the performance during the feedback phase.

In both groups, feedback was provided immediately after performing the first procedure. The total feedback time was held constant in both conditions and was approximately 10 minutes. Feedback was provided by 1 of the 2 researchers who were trained in providing both types of feedback in the context of this experiment. Care was taken to ensure that the feedback was complete and that participants were able to ask questions and verify whether they understood the information properly.

### Performance score

Video-recorded procedures were assessed blindly by 2 independent researchers using OSATS, as the most common assessment tool for surgical performance. A minimum of 1 and a maximum of 5 points could be awarded for each of the 7 competences. A maximum possible performance score for each procedure was 35 points. Both researchers were trained in assessment of recorded procedures. Training was facilitated by a surgeon who is an expert in this field. It included a comprehensive study of the procedure using the provided study material followed by execution of the procedure on the model themselves. After that, researchers were trained in assessment until they got sufficiently familiar with all aspects of OSATS. The actual assessment of recorded procedures was performed independently. In case of discrepancies, consensus was reached by re-evaluating the procedure. Additionally, 5% of procedures were randomly selected and assessed by the expert to detect any discrepancies in scoring. No differences in ratings were identified.

### Motion tracking

Motion tracking analysis was performed using a combination of a commercially available optical tracker system (PST Base, PS-Tech B.V., Amsterdam, The Netherlands) and a customized glove for the dominant right hand. This could track 6 degrees of freedom position in Cartesian coordinates (X, Y, and Z axis) at a rate of 30 samples per second. Time to complete the task and path length were measured. These have shown to be excellent markers of surgical performance.[11,36-38] Because not all participants were able to complete the procedure within 30 minutes, the completion of the step of hernia sac removal was chosen as the endpoint for the outcomes of motion tracking analysis.

### Outcomes

The study outcomes were defined as the differences in mean improvement in performance score (as measured by the OSATS assessment tool; time (in seconds) and path length (in meters) between the first and the second procedure between 2 groups. Outcomes were stratified by MRT scores. Individuals who scored below the mean were assigned to the MRT-low group (n = 22). Students who scored above the mean were assigned to the MRT-high group (n = 28).

## Statistical analysis

Because of the novelty of this study, no previous data were available to calculate the sample size. A sample size of 50 participants was assumed to be appropriate. Participants' baseline characteristics were summarized using descriptive statistics. Differences in baseline measurements were assessed with an independent *t* test for differences in means and $\chi^2$ test for differences in proportions. The differences in mean performance scores of the first procedure between groups were assessed with an independent *t* test. The improvement between second and first procedure within a group was assessed with a paired *t* test. The difference in mean improvement ($\Delta$) in performance scores between second and first procedure between groups were assessed with a 1-way ANCOVA. $\Delta$Performance score was included as dependent variable, intervention group and study time as fixed factor (0–1 vs 1–2 vs 2–3 hours), and performance score on the first procedure and MRT score as covariates. Additionally, the outcomes were stratified by MRT score to evaluate the effect of intervention for different levels of VSA. The analyses were repeated for mean improvement in time ($\Delta$ time) and path length ($\Delta$ path length). Partial eta squared was calculated and used as an effect size (0.2 = small effect, 0.5 = moderate effect 0.8 = large effect). Analyses were performed using SPSS statistical software package version 25.0 for Windows (IBM Corp, Armonk, NY). Statistical significance was determined at the level of $p < .05$.

## Results

A total of 50 medical students was included. There were no significant differences between groups on baseline characteristics, as shown in Table 1.

Both groups improved significantly in terms of total OSATS score, time, and path length between the first and second time of performing the procedure (Table 2). Since not all participants were able to complete the procedure within 30 minutes, the completion of the step of hernia sac removal was chosen as the endpoint for the outcome measures time (s) and path length (m). This step was performed by 42 (84%) of participants. Path length data of 5 of the participants was lacking due to technical issues.

The mean improvement in performance scores was not significantly different between the 2 groups ($\beta = 2.1$; 95% IC [-0.41 to -4.5]; $\eta2 = 0.06$; $p = .100$). However, the mean improvement in time ($\beta = -139.4$; 95% CI [-.262.5 to -16.5]; $\eta2 = 0.13$; $p = .027$) and in path length ($\beta = -21.2$; 95% CI [-41.9 to -0.5]; $\eta^2 = 0.13$; $p = .046$) was significantly greater in the OCHRA feedback group.

Table 1. Baseline characteristics of the included participants.

| | OCHRA feedback | OSATS feedback | p value |
|---|---|---|---|
| | n = 25 | n = 25 | |
| Sex, n (%) | | | |
| Male | 10 (40) | 13 (52) | .571 |
| Female | 15 (60) | 12 (48) | |
| Age, mean ± SD, in years | 21.5 ± 2.2 | 21.2 ± 1.9 | .537 |
| Study phase, n (%) | | | |
| Bachelor students | 15 (60) | 14 (56) | .302 |
| Master students | 10 (40) | 11 (44) | |
| Time spent studying online course, n (%) | | | |
| 0 – 1 hours | 8 (32) | 5 (20) | .288 |
| 1 – 2 hours | 16 (64) | 16 (64) | |
| 2 – 3 hours | 1 (4) | 4 (16) | |
| I liked the way the hernia repair was taught, median [IQR] | 8.0 [7.0-9.0] | 7.0 [6.5-8.7] | .104 |
| I felt prepared after completing the online course, mean ± SD | 6.3 ± 1.2 | 6.1 ± 2.1 | .679 |
| Times seen open inguinal hernia repair surgery in real life, median [IQR] | 0.0 [0.0-0.5] | 0.0 [0.0-0.5] | .984 |
| Other sources used to study, n (%) | | | |
| Not used | 16 (64) | 11 (44) | .256 |
| Yes | 9 (36) | 14 (56) | |
| Time spent studying other sources, n (%) | | | |
| 0 – 1 hour | 8 (88.9) | 13 (92.9) | 1.00 |
| 1 – 2 hours | 1 (11.1) | 1 (7.1) | |
| Hours of sleep last night, median [IQR] | 7.0 [6.0-8.0] | 8.0 [7.0-8.0] | .471 |
| Alcohol consumption last night, median [IQR] | 0.0 [0.0-0.8] | 0.0 [0.0- 0.0] | .402 |
| Coffee consumption before surgical performance, median [IQR] | 1.0 [0.0-1.0] | 0.5 [0.0-1.0] | .879 |
| Other circumstances that could have affected the surgical performance, n (%) | | | |
| Not used | 18 (72) | 22 (88) | .289 |
| Yes | 7 (28) | 3 (12) | |
| Mental Rotation Test score, mean ± SD | 16.4 ± 5.5 | 16.7 ± 4.9 | .872 |

OCHRA, Observational Clinical Human Reliability Analysis; OSATS, Objective Structured Assessment of Technical Skills; n, number of participants; SD, standard deviation; IQR, interquartile range.

Table 2. Differences in performance scores, time, and path length between 2 interventions.

| | OCHRA feedback | OSATS feedback | p value |
|---|---|---|---|
| **Performance score** | n = 25 | n = 25 | |
| 1st procedure | 17.4 ± 3.1 | 17.4 ± 3.8 | .935 |
| 2nd procedure | 23.5 ± 5.4 | 21.8 ± 4.9 | |
| Δ | 6.2 ± 3.5* | 4.4 ± 4.7* | .100[A] |
| **Time (sec)** | n = 20 | n = 22 | |
| 1st procedure | 1239.6 ± 274.8 | 1300.4 ± 382.3 | .561 |
| 2nd procedure | 868.6 ± 151.6 | 1025.7 ± 286.4 | |
| Δ | 371.0 ± 223.4* | 274.6 ± 341.6* | .027[A] |
| **Path length (m)** | n = 19 | n = 18 | |
| 1st procedure | 168.4 ± 61.5 | 168.9 ± 39.6 | .977 |
| 2nd procedure | 112.4 ± 36.2 | 134.2 ± 36.3 | |
| Δ | 53.5 ± 42.4* | 34.7 ± 39.0* | .046[A] |

Δ = delta, difference between 2nd and 1st procedure; sec, seconds; m, meters. *p < .001 paired t-test; [A] differences assessed with ANCOVA.

### Effect of VSA

When outcomes were stratified by MRT scores, the greater improvement in time in the OCHRA feedback group was observed only among individuals with lower VSA (β = -220.2; 95% CI [-418.4 to -22,1]; $\eta^2$ = 0.26; p = .032) (Figure 2).

As shown in Figure 3, a similar trajectory was observed for the improvement in path length. However, this difference did not reach the significance level (β = -28.2; 95% CI [-56.8 to 0.42]; $\eta^2$ = 0.24; p = .053). Regardless of intervention, MRT scores were significantly associated with mean improvement in time (β = -14.17; 95% CI [-26.9 to -2.6]; $\eta^2$ = 0.14; p = .019), but not in path length (β = -0.74; 95% CI [-2.8 to 1.3]; $\eta^2$ = 0.01; p = .469) and OSATS scores (β = 0.05; 95% CI [-0.2 to 0.3]; $\eta^2$ = 0.004; p = .670).

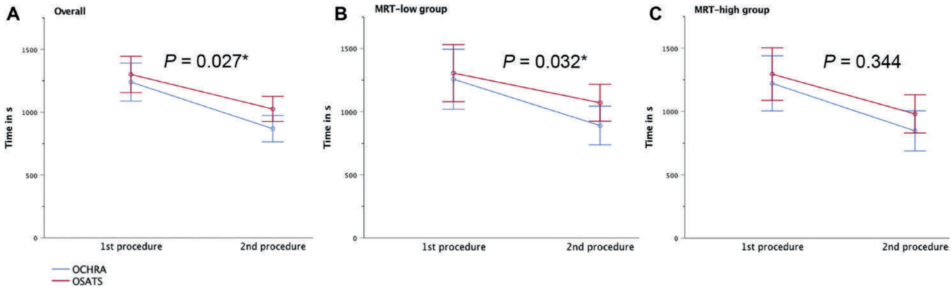


Figure 2. Differences in Δtime (s) between OCHRA feedback and OSATS feedback groups: (a) overall; (b) MRT-low group, and (c) MRT-high group; p < .05.

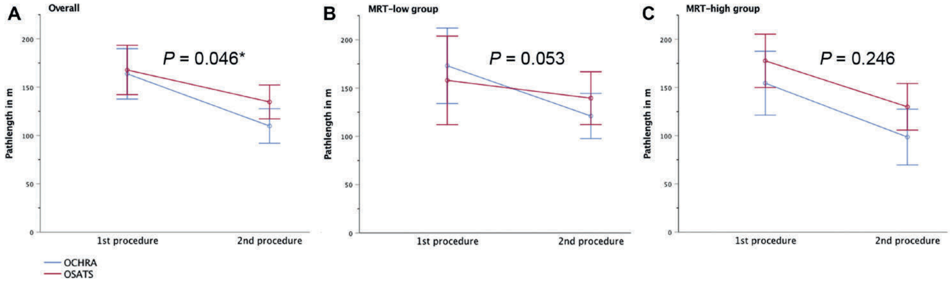


Figure 3. Differences in Δpath length (mm) between OCHRA feedback and OSATS feedback group: (a) overall; (b) MRT-low group, and (c) MRT-high group; $p < .05$.

# DISCUSSION

The aim of this study was to investigate whether a task-specific, stepwise feedback checklist (OCHRA) leads to a greater improvement in surgical performance compared to a global rating scale feedback method (OSATS). The outcomes were evaluated in relation to VSA. The mean improvement in performance scores was not significantly different between the OCHRA and OSATS feedback groups. However, the OCHRA feedback showed a significant improvement on performance in terms of time and path length, as measured by the hand-motion analysis system. The effects of OCHRA feedback were present mainly among individuals with lower VSA.

The observed effectiveness of OCHRA feedback on surgical performance in a simplified hernia repair model, as a more precise and concise approach, is supported by the instructional alignment theory.[39] When training and assessment methods are aligned, the effects of instruction are up to 4 times greater than in nonaligned methods.[39] In the current study, participants prepared for the open inguinal hernia repair procedure using a stepwise video demonstration. As OCHRA feedback was based on the evaluation of the subsequent surgical steps instead of competencies as part of the OSATS feedback, a greater alignment between learning and feedback could be achieved. Although this did not result in a difference in outcome in terms of the surgical scores, differences were found for the time and path length. In this study, most participants could not finish the entire surgical procedure within the 30-minute timeframe. Possibly, differences in surgical scores would have been found if students did complete the entire surgical procedure. Additionally, the value of a checklist (OCHRA) and global rating scale (OSATS) assessments may depend on the level of learners' experience.[40] Global rating scales have been reported to be more useful for learners with higher levels of expertise, whereas checklists may be more useful for novice learners, such as the participants in this study.[40,41]

The observed modifying effect of VSA on time and path length leads to important considerations. First, the findings are in line with previous research reporting positive association between VSA and hand motion.[15,42-45] However, by treating VSA as a possible effect modifier, this study showed that this association was present only for individuals with lower levels of VSA. This effect, also referred to as the aptitude-treatment effect,[46,47] has been repeatedly observed in the research field of anatomical education.[21-23,46] Therefore, it is instrumental to consider possible modifying effects of VSA on outcomes when designing new research. Second, the observed differences could be explained by the cognitive load theory.[48] Students with lower VSA are in general less effective in processing new spatial information in their working memory than students with higher VSA. However, in contrast to a global approach, the information from a task-specific stepwise feedback, building up on an already existing stepwise schema of a surgical procedure, could have decreased the cognitive load.[48] Subsequently, more working memory capacity could be created to process new procedural skills among low-performing individuals. This emphasizes the importance of an aptitude-based approach in learning and teaching surgical technical skills to novices. Lastly, the effect of VSA on OSATS scores was found to be not significant. This could be because of the inability of most participants to complete the entire procedure within the given timeframe.

OSATS was used both as an intervention and assessment scoring tool in this study. The rationale behind the choice to use the OSATS as an assessment scoring tool is that OSATS is considered to be the “gold standard” assessment tool for surgical performance and one of the few actually used in residency training and research.[5,15] In The Netherlands, OSATS is incorporated within the surgical residency training.[49] Second, a systematic review comparing checklists with global rating scales as assessment tools reported that global rating scales might be better in capturing nuanced elements of expertise.[40] Other assessment tools for surgical performance, such as the recently reported Surgical Quality Assurance (SQA), could have been an option, and perhaps would have found differences in surgical performance.[50]

The timing of feedback is still debated. Xeroulis et al distinguished feedback provided during the task (concurrent feedback) and feedback upon completing the task (summary feedback).[3] The latter was found to be superior for learning basic surgical skills; however, Al Fayyad et al found the opposite. In their study, concurrent (immediate) feedback was perceived as superior in learning basic surgical skills compared to summary (delayed) feedback.[51] In our study, summary feedback was chosen because the students operated on a simulation model without the risk of doing any harm. With an actual patient, a trainee needs guidance from a surgeon using concurrent feedback to avoid harmful errors.

### Limitations

This study has several limitations. First, the sample size could not be calculated beforehand due to the novelty of the study aim and design. Although it was sufficient to reveal significant differences in terms of time and path length, the sample size could have been too small to detect significant differences in OSATS scores. Second, not all participants were able to complete the procedure within given 30 minutes. As the step of hernia sac removal was reached by most participants, it was used as the endpoint to ensure a justified comparison in terms of time and path length. Allowing participants to complete the entire procedure would have provided a better display of their performance. Third, the participants were medical students with low and slightly various levels of anatomical knowledge and technical skills, including suturing. Due to random allocation, these differences are expected to have little to no effect on outcomes. Additionally, the mean improvement in outcome measures was chosen instead of the absolute scores to account for those differences. Another limitation is the possible inability to generalize the conclusions to left-handed students because this study only included right-handed students. Furthermore, these findings cannot be generalized to other procedures outside of inguinal hernia repair. Last, the effect of OCHRA feedback was evaluated in a simulated environment. This study should be repeated among surgical residents with higher levels of anatomical knowledge and technical skills in a clinical setting on multiple procedures.

The findings of this study have implications for both practice and research. In this study, the open inguinal hernia repair was chosen as an exemplary procedure. It is unknown whether an inguinal hernia repair simulation is ideally suited to detect the effects of different types of feedback on study outcomes. The implementation of structured, stepwise feedback that is aligned with the learning activities should be considered especially in the early phases of surgical training. The aligned stepwise instruction using stepwise video demonstrations and procedure specific OCHRA checklist assessment can be transferred to other surgical procedures. The stepwise segmentation of a surgical procedure can be made using the step-by-step framework.[26] This stepwise description of a surgical procedure can then be used to create a procedure-specific OCHRA checklist. Moreover, an aptitude-based approach in teaching and learning of surgical procedural skills could be of benefit for individuals with lower VSA. As demonstrated, it is crucial to consider the modifying effect of VSA on surgical outcomes when setting up new research. In fact, when overall outcomes are not evaluated for different levels of VSA, the real differences may remain unrevealed.

# CONLUSIONS

In conclusion, a task-specific, stepwise feedback checklist (OCHRA) proves to be more effective in improving surgical skills, in terms of time and path length, among surgical novices compared to a global rating scale feedback (OSATS). The effects of a task-specific feedback are present mostly in individuals with lower levels of VSA.

# REFERENCES

1. El Boghdady M, Alijani A. Feedback in surgical education. The Surgeon 2017;15(2):98-103.
2. Porte MC, Xeroulis G, Reznick RK, Dubrowski A. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. Am J Surg. 2007;193(1):105-10.
3. Xeroulis GJ, Park J, Moulton C-A, Reznick RK, LeBlanc V, Dubrowski A. Teaching suturing and knot-tying skills to medical students: A randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. Surgery 2007;141(4):442-9.
4. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. Am J Surg. 2011;202(4):469-80.
5. Van Hove P, Tuijthof G, Verdaasdonk E, Stassen L, Dankelman J. Objective assessment of technical surgical skills. BJS 2010;97(7):972-87.
6. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. BJS 1997;84(2):273-8.
7. Niitsu H, Hirabayashi N, Yoshimitsu M, Mimura T, Taomoto J, Sugiyama Y, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. Surgery today. 2013;43(3):271-5.
8. Tang B, Hanna GB, Joice P, Cuschieri A. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. Arch Surg. 2004;139(11):1215-20.
9. Biggs J. Enhancing teaching through constructive alignment. Higher educ. 1996;32(3):347-64.
10. Levin M, McKechnie T, Khalid S, Grantcharov TP, Goldenberg M. Automated Methods of Technical Skill Assessment in Surgery: A Systematic Review. J Surg Educ 2019;76(6):1629-39.
11. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg. 2001;193(5):479-85.
12. Shaharan S, Nugent E, Ryan DM, Traynor O, Neary P, Buckley D. Basic Surgical Skill Retention: Can Patriot Motion Tracking System Provide an Objective Measurement for it? J Surg Educ 2016;73(2):245-9.
13. Genovese B, Yin S, Sareh S, DeVirgilio M, Mukdad L, Davis J, et al. Surgical Hand Tracking in Open Surgery Using a Versatile Motion Sensing System: Are We There Yet? The American surgeon. 2016;82(10):872-5.
14. Corvetto MA, Fuentes C, Araneda A, Achurra P, Miranda P, Viviani P, et al. Validation of the imperial college surgical assessment device for spinal anesthesia. BMC anesthesiology. 2017;17(1):131.
15. 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(5):750-7.
16. Gordon HW. The cognitive laterality battery: tests of specialized cognitive function. Int. J. Neurosci. 1986;29(3-4):223-44.
17. Kozhevnikov M, Hegarty M. A dissociation between object manipulation spatial ability and spatial orientation ability. Memory & cognition. 2001;29(5):745-56.

18. Vajsbaher T, Schultheis H, Francis NK. Spatial cognition in minimally invasive surgery: a systematic review. BMC surgery. 2018;18(1):94.
19. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and technical skills performance in health care: a systematic review. Med Educ 2015;49(11):1065-85.
20. Roach VA, Brandt MG, Moore CC, Wilson TD. Is three-dimensional videography the cutting edge of surgical skill acquisition? Anat Sci Educ 2012;5(3):138-45
21. Bogomolova K, van der Ham IJM, Dankbaar MEW, van den Broek WW, Hovius SER, van der Hage JA, et al. The Effect of Stereoscopic Augmented Reality Visualization on Learning Anatomy and the Modifying Effect of Visual-Spatial Abilities: A Double-Center Randomized Controlled Trial. Anat Sci Educ 2020;13:558-567
22. 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(1):34-45.
23. Roach VA, Mistry MR, Wilson TD. Spatial visualization ability and laparoscopic skills in novice learners: evaluating stereoscopic versus monoscopic visualizations. Anat Sci Educ 2014;7(4):295-301.
24. Tchantchaleishvili V, Myers PO. Left-handedness—a handicap for training in surgery? J Surg Educ 2010;67(4):233-6.
25. Nazari T, Simons M, Zeb M, van Merriënboer J, Lange J, Wiggers T, et al. Validity of a low-cost Lichtenstein open inguinal hernia repair simulation model for surgical training. Hernia 2020:24:895-901.
26. Nazari T, Vlieger E, Dankbaar M, van Merriënboer J, Lange J, Wiggers T. Creation of a universal language for surgical procedures using the step-by-step framework. BJS Open. 2018;2(3):151-7.
27. Nazari T, van de Graaf FW, Dankbaar MEW, Lange JF, van Merriënboer JJG, Wiggers T. One Step at a Time: Step by Step Versus Continuous Video-Based Learning to Prepare Medical Students for Performing Surgical Procedures. J Surg Educ 2020;77(4):779-87.
28. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ 2017;10(3):235-41.
29. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. Perceptual and motor skills 1978;47(2):599-604.
30. Wanzel KR, Hamstra SJ, Anastakis DJ, Matsumoto ED, Cusimano MD. Effect of visual-spatial ability on learning of spatially-complex surgical skills. Lancet 2002;359(9302):230-1.
31. Foster J, Miskovic D, Allison A, Conti J, Ockrim J, Cooper E, et al. Application of objective clinical human reliability analysis (OCHRA) in assessment of technical performance in laparoscopic rectal cancer surgery. Techniques in coloproctology. 2016;20(6):361-7.
32. Miskovic D, Ni M, Wyles SM, Parvaiz A, Hanna GB. Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. Surgical endoscopy. 2012;26(3):796-803.
33. Talebpour M, Alijani A, Hanna G, Moosa Z, Tang B, Cuschieri A. Proficiency-gain curve for an advanced laparoscopic procedure defined by observation clinical human reliability assessment (OCHRA). Surg Endosc 2009;23(4):869-75.
34. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. Academic medicine: journal of the Association of American Medical Colleges. 1996;71(12):1363-5.
35. Vaidya A, Aydin A, Ridgley J, Raison N, Dasgupta P, Ahmed K. Current Status of Technical Skills Assessment Tools in Surgery: A Systematic Review. J Surg Research. 2020;246:342-78.

36. Atesok K, Satava RM, Marsh JL, Hurwitz SR. Measuring surgical skills in simulation-based training. JAAOS-Journal of the American Academy of Orthopaedic Surgeons. 2017;25(10):665-72.
37. Binkley J, Bukoski AD, Doty J, Crane M, Barnes SL, Quick JA. Surgical simulation: markers of proficiency. J Surg Educ 2019;76(1):234-41.
38. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. J Am Surg 2002;184(1):70-3.
39. Cohen SA. Instructional alignment: Searching for a magic bullet. Educational Researcher. 1987;16(8):16-20.
40. Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ 2015;49(2):161-73.
41. Dankbaar ME, Stegers-Jager KM, Baarveld F, van Merrienboer JJ, Norman GR, Rutten FL, et al. Assessing the assessment in emergency care training. PLoS One. 2014;9(12):e114663.
42. Groenier M, Schraagen JM, Miedema HA, Broeders IA. The role of cognitive abilities in laparoscopic simulator training. Adv Health Sci Educ 2014;19(2):203-17.
43. Hedman L, Ström P, Andersson P, Kjellin A, Wredmark T, Felländer-Tsai L. High-level visual-spatial ability for novices correlates with performance in a visual-spatial complex surgical simulator task. Surg Endosc 2006;20(8):1275-80.
44. Schlickum M, Hedman L, Enochsson L, Henningsohn L, Kjellin A, Felländer-Tsai L. Surgical simulation tasks challenge visual working memory and visual-spatial ability differently. World J Surg 2011;35(4):710-5.
45. Hedman L, Klingberg T, Enochsson L, Kjellin A, Felländer-Tsai L. Visual working memory influences the performance in virtual image-guided surgical intervention. Surg Endosc 2007;21(11):2044-50.
46. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med 2005;80(6):541-8.
47. Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandenbroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. Clin Epidemiol 2017;9:331-8.
48. van Merrienboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ 2010;44(1):85-93.
49. Hopmans CJ, den Hoed PT, van der Laan L, van der Harst E, van der Elst M, Mannaerts GH, et al. Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. Surgery 2014;156(5):1078-88.
50. Tsai AY, Mavroveli S, Miskovic D, van Oostendorp S, Adamina M, Hompes R, et al. Surgical Quality Assurance in COLOR III: Standardization and Competency Assessment in a Randomized Controlled Trial. Ann Surg 2019;270(5):768-74.
51. Al Fayyadh MJ, Hassan RA, Tran ZK, Kempenich JW, Bunegin L, Dent DL, et al. Immediate auditory feedback is superior to other types of feedback for basic surgical skills acquisition. J Surg Educ 2017;74(6):e55-e61.

# 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.

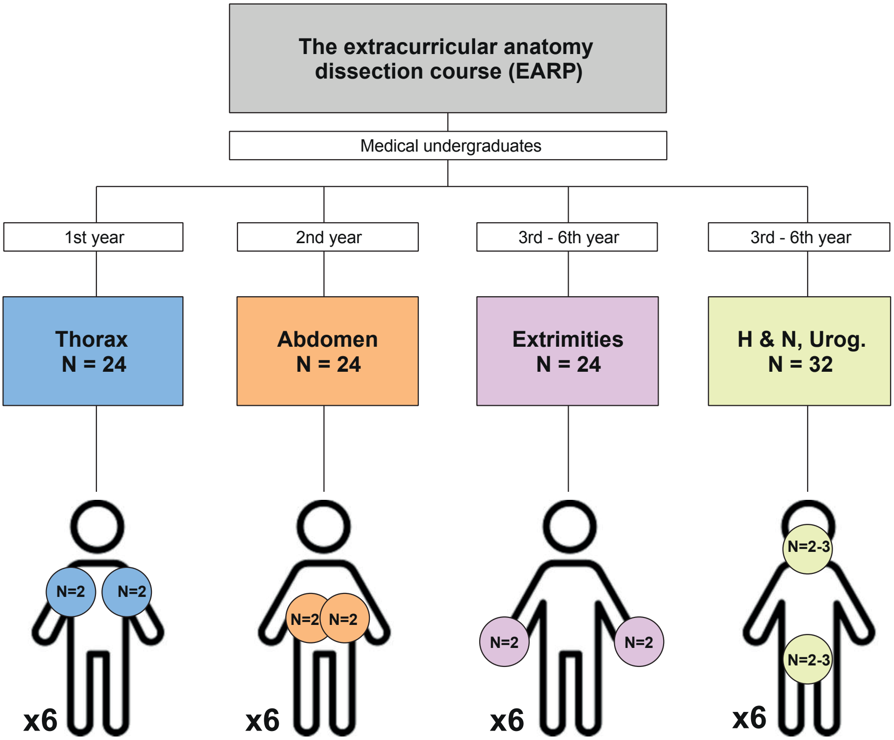


Figure 1. An extra-curricular anatomy dissection course. Students attend eight dissection sessions of three hours each week. EARP, the Erasmus MC Anatomy Research Project; H & N, Urog., Head & Neck and Urogenital System.

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._

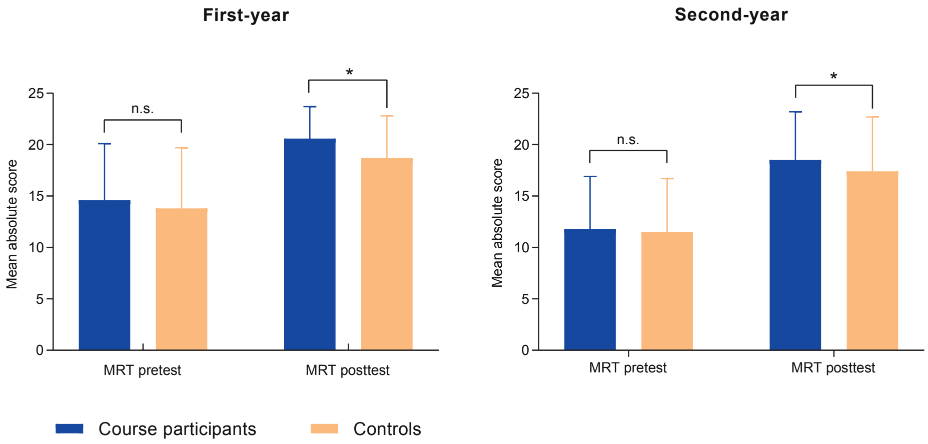


Figure 2. Differences in performance on MRT pretest and posttest between course participants (*first year*: EARP Thorax program, *second year*: EARP Abdomen program) and controls. Performances are reported in mean scores. Error bars represent standard deviation; *p < .05; n.s., not significant; MRT, 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%, $\chi^2$ = 3.091, p = .079).

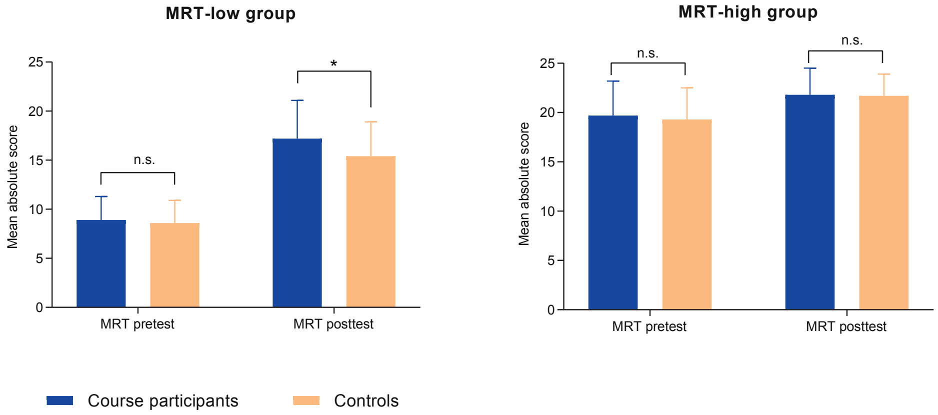


Figure 3. Differences in performance on MRT pretest and posttest between course participants and controls in the MRT-low and MRT-high groups of first- and second-year medical undergraduates. Performances are reported in mean scores. Error bars represent standard deviation; *p < .05; n.s., not significant;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

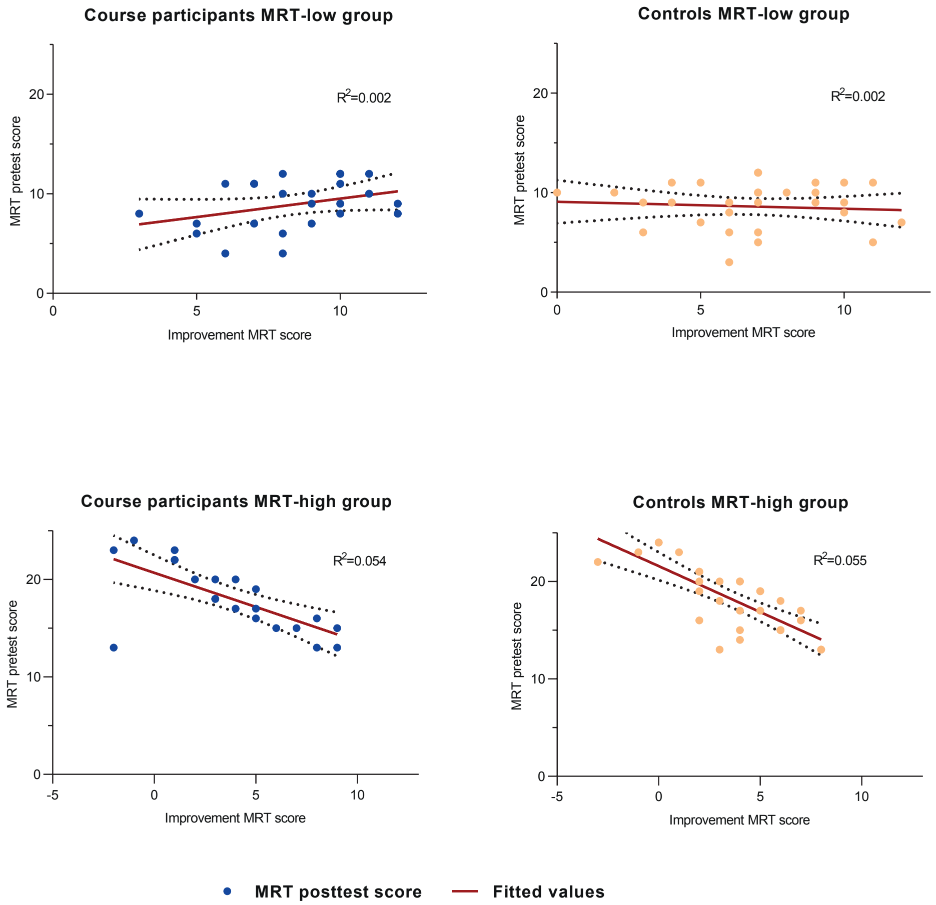


Figure 4. Relationship between MRT pretest scores and mean improvement in the MRT scores. A regression analysis graph illustrating the relationship between initially low and high levels of visual-spatial abilities and mean improvement among course participants and controls. MRT, Mental Rotations Test.

# 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.

# 8

# General discussion and future perspectives

# GENERAL DISCUSSION AND FUTURE PERSPECTIVES

Currently, medical students mostly learn from two-dimensional (2D) images in textbooks and anatomical atlases, and they experience difficulties in translating acquired anatomical knowledge into practice. The latter is greatly affected by the level of visual-spatial abilities (VSA), i.e., the ability to construct visual-spatial, i.e., three-dimensional (3D) mental representations of 2D images and to mentally manipulate these representations.[1,2] VSA significantly predicts the level of anatomical knowledge among students and surgical performance among residents in the early phases of surgical training.[3,4]

Three-dimensional visualization technology (3DVT) has the potential to fill the gap between learning and applying anatomical knowledge in practice. The great advantage of 3DVT is its ability to visualize anatomical structures and explore spatial relations between various structures from numerous viewpoints and angles. One of the distinguishing features of new generations of 3DVT is their ability to provide accurate depth perception or stereoscopic vision. Stereoscopic vision results from the binocular disparity that various supportive devices can obtain by projecting a slightly different image to the left and right eyes. In this thesis, this type of technology is referred to as *stereoscopic 3DVT*. Examples of stereoscopic 3DVT include stereoscopic 3D desktop, stereoscopic 3D augmented reality (AR), and virtual reality (VR) environments (*see Figure 3, Introduction*). In the absence of stereoscopic vision, depth perception is created by so-called monocular depth cues that require only one eye for perception. They include cues such as coloring, shading, and motion parallax.[5] In this thesis, this type of technology is referred to as *monoscopic 3DVT.* Examples of traditional monoscopic 3DVT include 3D anatomical models viewed on a 2D computer screen (*see Figure 3, Introduction*).

The overarching aim of this thesis was to evaluate the effectiveness of stereoscopic 3DVT and determine how various levels of VSA would interact with learning using this technology to improve anatomical and surgical education. The predictions in this thesis were made based on the model of Cognitive Load Theory (CLT).[6] This theory assumes that human working memory, which processes new information, has limited capacity. The working memory load consists of three sources: intrinsic cognitive load (nature of learning material content), extraneous cognitive load (the way learning material is presented), and germane cognitive load (the actual learning process).[7] When the sum of three sources exceeds the working memory capacity, cognitive overload occurs, impairing learning. Based on this theory, individuals with lower levels of VSA devote more available cognitive resources to mental visualization and manipulation of 3D objects than individuals with higher levels of VSA. Consequently, when learning spatially-complex material from 2D images, low-VSA individuals are left with fewer resources to spend on actual learning tasks. A potential way to facilitate learning is to decrease the input of other sources of cognitive load by improving the instructional design or changing the way learning material is presented such that learners will be able to re-allocate their resources to actual learning. Within this framework, providing stereoscopic vision in 3DVT can be interpreted as improving instructional methods to improve learning.

In this chapter, the main findings are discussed in light of CLT, and educational implications and suggestions for future directions are provided.

### The role of stereoscopic vision in learning with 3DVT

As previously described, the '3D effect', or visual depth perception, in 3DVT is shaped by a mental combination of monocular and binocular depth cues. In this thesis, we evaluated the role of stereoscopic vision (binocular cue) as one of the most substantial providers of visual depth perception in 3DVT. By performing a meta-analysis, we demonstrated that providing stereoscopic vision has a significant positive effect on learning anatomy with 3D anatomical models. This finding emphasizes the importance of distinguishing between monoscopic and stereoscopic 3DVT **(chapter 2)**.

In the context of CLT, these differences are explained by two mental processes involved in learning with monoscopic and stereoscopic visualizations. A digital 3D object viewed monoscopically is memorized as a set of screenshots, or so-called key view-based 2D images.[8,9] These 2D images are then mentally combined to reconstruct a total mental representation of the 3D object. Consequently, this process consumes a relatively large amount of working memory capacity and leaves fewer cognitive resources for actual learning. In stereoscopic visualization, by contrast, mental representation of the 3D object is already built and provided by stereoscopic vision. The mental reconstruction required in monoscopic visualization can be skipped in stereoscopic visualization while leaving sufficient cognitive resources for other learning tasks. In other words, these findings strongly suggest that mental representations do not primarily consist of key view-based 2D images; instead, they might also include spatial information depending on the type of input. While monoscopic visualization stimulates mental key view-based 2D images, stereoscopic visualization stimulates structural 3D mental representations. This conclusion is further supported by neurocognitive research. Binocular cues appear to activate neurons in the brain that differ from those activated by monocular cues.[10-12] Researchers demonstrated using electroencephalography that learning with stereoscopic 3D models resulted in greater 3D object recognition than obtained using monoscopic 3D models.[13]

As was demonstrated in our follow-up studies **(chapters 4 and 5)**, the effect of stereoscopic vision was not comparable for all types of 3D environments. Our findings suggest that cues such as motion parallax provide sufficient visual depth perception, depending on the type of 3D technology used**.** The reasoning is based on the results of our study evaluating the effect of stereoscopic vision in a 3D AR environment **(chapter 4).** In that study, the stereoscopic vision did not provide better learning effects. However, in that study, students could walk around the model and explore it from several possible angles. According to research in neurocognitive sciences, motion parallax, in some cases, can provide even more effective depth cues than stereoscopic vision alone.[14-17] Furthermore, motion parallax compensates for the absence of stereoscopic vision and improves the recognition of 3D shapes.[18] It remains unknown to what extent stereoscopic vision and motion parallax (combined or separately) contribute to visual depth perception in a stereoscopic 3D AR environment. Recent research suggests that the effect of stereoscopic vision varies among types of technology. Wainman and colleagues performed a similar study to evaluate the effect of stereoscopic vision in a 3D AR environment.[19] They also found that stereoscopic vision in a 3D AR environment did not contribute to learning as expected. Additionally, authors compared the effect of stereoscopic vision in a 3D AR environment with its effect in VR. The effect of stereoscopic vision in VR appeared to be significantly greater than in AR. Unfortunately, due to stereoscopic 3D AR technology's novelty, no other studies are available to compare stereoscopic 3D AR technology's effectiveness with others. Future research will provide more clarity on this subject.

### Aptitude-treatment interaction caused by VSA

Supported by the evidence of a positive effect of stereoscopic vision, we performed follow-up studies to evaluate the effects of VSA on learning with stereoscopic 3DVT. In the context of CLT, we hypothesized that providing stereoscopic vision would improve learning and that the most significant effect would be observed among individuals with low VSA. As demonstrated in **chapters 3, 5, and 6**, VSA greatly affected learning outcomes, although in different directions.

First, VSA caused an **aptitude-treatment interaction (ATI).** This interaction occurs when an individual with characteristic 1 learns better with instructional method A than with method B, while an individual with characteristic 2 learns better with method B.[20] In this thesis, ATI was observed in learning anatomy among students **(chapter 3)** and surgical procedures among residents **(chapter 5)** using stereoscopic 3DVT. Second, the observed ATI was not consistent across studies. VSA appeared to behave in two different directions, explained by two co-existing mechanisms, or hypotheses, in the context of CLT.[21]

*VSA as compensator*

The ability-as-compensator hypothesis predicts that individuals with high VSA can compensate for poor (monoscopic) instructional methods and, simultaneously, that improved (stereoscopic) instructional methods *compensate* for the lack of resources in individuals with low VSA. In other words, in the context of stereoscopic visualization, stereoscopic vision can act as a 'cognitive prosthetic' and improve learning in individuals with low VSA.[22] This mechanism was demonstrated in **chapter 3**. Learning with a stereoscopic 3D AR model was more effective than learning with a monoscopic 3D model only for students with low VSA. High VSA students performed equally well in all conditions.

*VSA as enhancer*

The ability-as-enhancer hypothesis predicts that high levels of VSA are required to derive benefit from the improved instructional method, while individuals with low levels of VSA are hindered due to increased demands imposed by processing new information.[23,] This mechanism was demonstrated in **chapter 5,** where surgical residents performed a spatially-complex procedure on a simulation model after watching an instructional video of the procedure either in 2D (monoscopically) or 3D (stereoscopically with active 3D glasses). Within this context, only residents with high VSA benefited from stereoscopic visualization and performed the procedure significantly better than with monoscopic visualizations. As novices with no prior knowledge, residents with low VSA were probably hindered by the high degree of visual interactivity in the instructional video and could not allocate sufficient resources to benefit from stereoscopic visualization.

In the context of monoscopic visualization, the enhancing mechanism explains why students with high VSA benefited from learning with monoscopic 3D models while low VSA students could not. This mechanism was recognized by Garg and colleagues more than two centuries ago.[25-27] Later, Huk tested and confirmed the hypothesis by evaluating the performance and the perceived cognitive load of medical students learning cell biology with monoscopic 3D desktop models.[28] Students with low VSA performed significantly worse on the test and reported their cognitive load to be high, whereas the opposite was observed for students with high VSA.

While *compensating* mechanisms in learning with stereoscopic 3DVT are recognized in anatomical education research[29-31], the *enhancing* mechanism has not been described previously. In other fields, including educational psychology and science, technology, engineering, and mathematics (STEM), the dichotomy between the two mechanisms has been widely recognized.[21,32] The majority of research in these disciplines has been performed in learning with multi-media. Similar to our findings, the observed mechanism depended on the type of instructional method and level of VSA. In addition to VSA, prior knowledge has also been recognized as a direct modifier.[22] Together with the existing body of evidence, our findings imply that learning with stereoscopic 3DVT can be effective for both low- and high-VSA individuals depending on their level of prior knowledge and expertise. This realization also means that for low-VSA individuals to benefit from stereoscopic visualization, learning material content should not be too complex or should build on prior knowledge.

### Personalized approach with VSA

The demonstrated ATI caused by VSA in learning with 3DVT (**chapters 3 and 5)** highlighted the importance of an individualized approach in medical training. In **chapter 6**, we demonstrated that an individualized approach based on VSA can make a difference. In a randomized controlled trial, we compared two types of intraoperative feedback on the performance of a spatially-complex procedure by medical undergraduates. Only after the results were stratified by VSA the fundamental differences were revealed. Students with low VSA performed significantly better after receiving task-specific, stepwise feedback than after receiving global rating scale feedback. Students with high VSA, however, benefited from both types of feedback. The findings can be relevant for both medical and higher education in general. They emphasize the importance of congruence between learning and assessment, as proposed by the constructive alignment theory[33], and between learning and feedback. This congruence provides opportunities for teachers who provide feedback or monitor peer feedback to evaluate the extent of this alignment on an ongoing basis and re-align them when needed. Feedback should be scaffolded with the learning activities, outcomes, and assessment.[34]

### The malleability of VSA

Another important finding of this thesis was that VSA, regardless of intervention type, was significantly associated with anatomical knowledge (**chapters 3 and 4)** and surgical performance **(chapter 5).** More importantly, this association was not comparable for all levels of VSA. As was shown in **chapter 5**, only low levels of VSA were positively associated with surgical performance. Among the studies reporting associations between VSA and surgical performance, only one study recognized a similar pattern by considering the possible ATI caused by VSA.[35] These findings suggest that VSA affects surgical performance when it is not well developed. They also suggest that when VSA is well developed, surgical performance is predicted by factors such as surgical experience. If VSA appears to be such an essential factor in learning anatomy and performing surgical procedures, the following questions arise: can VSA be trained or be improved by repeated practice, starting from the early stages of medical training?

In **chapter 7**, we demonstrated that repeated practice of cadaveric dissections improved VSA of medical undergraduates. The improvement of VSA occurred only among students with initially low levels of VSA, again reflecting the ATI caused by VSA. It is needed to say that cadaveric dissections are not the only effective way of training VSA. Methods that stimulate mental visualization and manipulation, such as evaluating cross-sections and mental rotation training, also effectively improve VSA.[36] As demonstrated in STEM domains, VSA training is practical, durable, and transferable across all categories of spatial skills.[37] The findings that VSA can be improved by training opens a new window of opportunity. It means that by improving VSA, one can improve its own level anatomical knowledge and surgical performance.

## Reflections on the methodology

The merits of the results of this thesis are interconnected with its strengths and limitations. The overall methodological rigor is reflected in the use of experimental study designs that are essential to understand the working mechanisms of 3DVT. To avoid confounders - not uncommon in media-comparative research - most comparisons in this thesis were made within a single level of instructional design.[20] In this way, we examined the effect of stereoscopic vision as the only truly manipulated element in study designs. Additionally, to increase our findings' generalizability and share resources among collaborative institutes, we conducted multiple multicenter studies.

One of the limiting factors concerns using the Mental Rotation Test (MRT) as the assessment tool for VSA across studies. First, this test measures mental rotation as one of the two components of VSA. As the other component of VSA, mental visualization has been measured by the Paper Folding Test (PFT) but did not show any associations with our outcomes. Despite the demonstrated ability of PFT to measure mental visualization and transformation, other validated psychometric tests exist that could have impacted our findings if used. These tests include the Embedded Figure Test and the Mental Cutting Test, which are widely used in educational psychology and STEM.[37] Likewise, various other psychometric tests for measuring mental rotation are available. These include the Card Rotation Test, the Cube Comparison Test, and the Purdue Spatial Visualization Test.[37] Perhaps using the term 'mental rotation skill' would be a more appropriate term for VSA in this thesis.

Another interesting detail regarding the use of MRT is the difference in performance between sexes. Sex differences in MRT scores have been repeatedly reported in the literature and were observed in our studies as well.[37-41] However, evidence suggests that these differences are not primarily caused by actual differences in working memory capacity but can be altered by chosen strategy, confidence, and familiarity with the presented 3D objects. Research has shown that other strategies can be used to solve mental rotation tasks.[42] While men often use global-shape strategies, women often choose to analyze local aspects of the figure shapes.[43] The use of the global-shape strategy is associated with better test performance, suggesting that choice of strategy can give rise to sex differences. The differences in cerebral activation patterns further support the notion that women use different strategies to solve mental rotation tasks despite equal performance on the Mental Rotation Test.[44] In addition, females tend to rethink their choices, causing them to complete fewer items than men,[45,46] who rely on a 'leaping' strategy by moving on to the next item as soon as they have identified the answer.[47] Lastly, familiarity with manipulated 3D objects refers to the type of stimulus that can cause sex differences. The latter is supported by various studies in which human body parts were used instead of traditional cubes figures, minimizing the notable sex differences on performance.[48,49] Taken together, this could mean that the sex differences we found in our studies, and those in the literature, may have produced a distorted view of reality. In other words, the actual levels of VSA among females may have been underestimated by using an inappropriate instrument.

## Future implications

The findings of this thesis lead to several important considerations for educational research and practice.

### *An individualized approach in learning with 3DVT*

One fact emerges regarding the educational role of 3DVT in the medical curriculum: one size does not fit all. Individual learning needs of students should be paramount when determining whether and how 3D technology can be implemented. As demonstrated in this thesis, stereoscopic visualization can benefit both low- and high-VSA learners, depending on the complexity of the learning content and learners' prior knowledge. For example, monoscopic 3D models should be avoided for low VSA students because monoscopic visualization comprises learning by inducing cognitive overload. Instead, stereoscopic 3D models can be considered an additional teaching tool for low-VSA students. Small group sessions can be considered for using stereoscopic 3D AR technology because it enables collaboration and active learning simultaneously. In surgical procedure training, where the complexity of learning content increases, stereoscopic visualization of instructional videos is recommended for high-VSA residents only. For residents with low VSA, traditional 2D videos would be sufficient.

Future research should focus on the working mechanisms of 3DVT in anatomical and surgical education. This research will eventually aid the implementation of suitable types of technology in the correct educational settings. For example, future studies could focus on the effect of motion parallax in 3D AR environments on learning anatomy with or without stereoscopic vision. When essential features or elements of stereoscopic 3D AR technology are known, they can be effectively incorporated into instructional activities for learning with 3DVT. It would also be helpful to evaluate these effects concerning objectively measured cognitive loads. This evaluation can be performed by integrating eye-tracking functions within the 3D AR devices. Care should be taken to moderate factors such as prior knowledge, expertise, and complexity of the learning material.

*Accounting for ATI caused by VSA*

For research purposes, it is essential to emphasize the role of the aptitude-treatment interaction (ATI) caused by VSA on learning. In statistical terms, this phenomenon is called 'effect measure modification'.[50,51] As demonstrated in our studies **(chapters 2, 3, 5, and 6)**, this interaction can be revealed either by stratifying the overall results by VSA or by including VSA as an interaction term in the regression analysis. Adjusting for VSA only as a confounder will not reveal this interaction. Although ATI is widely recognized in educational psychology and STEM domains, it has been hardly mentioned in anatomical and surgical education research. Therefore, it is essential to account for the potential aptitude-treatment effect of VSA when planning new studies and analyzing the data.

*VSA as an identification tool*

Until now, VSA has been seen as a fixed individual characteristic that does not change over time. It is not surprising that VSA has been recommended as a selection tool in surgical training. However, the malleability of VSA demonstrated in our studies strongly suggests that VSA can be trained and improved. Therefore, one should consider using VSA as an identification tool rather than a selection tool. The goal is to identify individuals with low levels of VSA and provide them with practical tools for both VSA and anatomical and surgical training. This goal will be of great interest for students who pursue surgical careers but whose VSA skills need improvement.

Future research should focus on exploring the optimal ways of improving low levels of VSA and implementing VSA training in medical curricula. Using 3D technology could also be a training tool along with practical (dissection-based) and theoretical (mental rotation exercises) methods. The emphasis should be placed on the elements that are essential for the actual improvement of VSA. Simultaneously, the effect of VSA training on anatomical knowledge and surgical performance should be evaluated. By understanding the building blocks of practical VSA training and their effect on learning, various methods can be effectively designed and implemented. Next to improving low levels of VSA, it would be interesting to explore the effect of VSA training on high-performing individuals. Would it be possible for these individuals to improve their VSA further and achieve even higher performance levels? In other words, can we help excellent students excel even further? For this part of research, VSA assessment instruments other than the gold standard should be considered to avoid the possible ceiling effect among high-performance individuals.

# REFERENCES

1. Gordon HW. The cognitive laterality battery: tests of specialized cognitive function. *Int J Neurosci.* 1986;29(3-4):223-244.
2. Kozhevnikov M, Hegarty M. A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem Cognit.* 2001;29(5):745-756.
3. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. Anat Sci Educ 2017;10:235–241.
4. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and technical skills performance in health care: a systematic review. *Med Educ.* 2015;49(11):1065-1085.
5. Johnston EB, Cumming BG, Landy MS. Integration of stereopsis and motion shape cues. Vision Res. 1994;34(17):2259-2275.
6. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. 2010;44(1):85-93
7. Sweller J, van Merrienboer JJG, Paas F. Cognitive architecture and instructional design: 20 years later. *Educ Psychol Rev.* 2019;31(2):261-292.
8. Garg AX, Norman G, Sperotable L. How medical students learn spatial anatomy. Lancet. 2001;357(9253):363-364.
9. Bulthoff HH, Edelman SY, Tarr MJ. How are three-dimensional objects represented in the brain? Cereb Cortex. 1995;5(3):247-260.
10. Jolicoeur P, Milliken B. Identification of disoriented objects: effects of context of prior presentation. J Exp Psychol Learn Mem Cogn 1989;15:200–210.
11. Kourtzi Z, Erb M, Grodd W, Bulthoff HH. Representation of the perceived 3-D object shape in the human lateral occipital complex. Cereb Cortex 2003;13:911–920.
12. Verhoef B-E, Vogels R, Janssen P. Binocular depth processing in the ventral visual pathway. Philos Trans R Soc Lond B Biol Sci 2016;371:20150259.
13. Anderson SJ, Jamniczky HA, Krigolson OE, Coderre SP, Hecker KG. Quantifying two-dimensional and three-dimensional stereoscopic learning in anatomy using electroencephalography. NPJ Sci Learn. 2019;4:10.
14. Bradshaw MF and Rogers BJ. The interaction of binocular disparity and motion parallax in the computation of depth. Vision Res. 1996; 21:3457-3468.
15. Naepflin U and Menozzi M. Can movement parallax compensate lacking stereopsis in spatial explorative search tasks? Displays 2001; 22(5): 157-164.
16. Aygar E, Ware C, Rogers D. The Contribution of Stereoscopic and Motion Depth Cues to the Perception of Structures in 3D Point Clouds. ACM Trans. Appl. Percept. 2018; 15(2):1-13
17. Lankheet MJM and Palmen M. Stereoscopic segregation of transparent surfaces and the effect of motion contrast. Vision Research 1998; 38(5):569-668
18. Sherman A, Papathomas TV, Jain A, Keane BP. The role of stereopsis, motion parallax, perspective and angle polarity in perceiving 3-D shape. Seeing Perceiving 2012;25:263–285.
19. Wainman B, Pukas G, Wolak L, Mohanraj S, Lamb J, Norman GR. The critical role of stereopsis in virtual and mixed reality learning environments. Anat Sci Educ 2020;13:398–405
20. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. Acad Med. 2005;80(6):541-548.
21. Mayer R and Sims VK. Whom Is a Picture Worth a Thousand Words?: Extensions of a Dual-Coding Theory of Multimedia Learning. J Educ Psychol 1994; 86(3):389-401

22. Höffler TN. Spatial ability: Its influence on learning with visualizations – a meta-analytic review. Educ Psychol Rev 2010; 22:245-269
23. Hegarty M, Sims VK. Individual differences in use of diagrams as external memory in mechanical reasoning. Mem Cognit *1994;* 22:411–430.
24. Mayer RE. Multimedia Learning. *3rd Ed.* Cambridge, UK*: Cambridge University Press. 2001. 452 p.*
25. Garg A, Norman GR, Spero L, Maheshwari P.. Do virtual computer models hinder anatomy learning? Acad Med 1999; 74:S87–S89.
26. Garg A, Norman G, Spero L, Taylor I. Learning anatomy: Do new computer models improve spatial understanding? Med Teach 1999; 21:519–522.
27. 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.
28. Huk T. Who benefits from learning with 3D models? The case of spatial ability. J Comput Assist Learn 2006; 22:392–404.
29. 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.
30. Luursema JM, Verwey WB, Kommers PA, Geelkerken RH, Vos HJ. Optimizing conditions for computer-assisted anatomical learning. Interact Comput 2006; 18:1123–1138.
31. Luursema JM, Verwey WB, Kommers PA, Annema JH. The role of stereopsis in virtual anatomical learning. Interact Comput 2008; 20:455–460.
32. Barrett TJ and Hegart M. Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. Comp Human Behav 2016; 65:220-231
33. Biggs J. Enhancing teaching through constructive alignment. High Educ 1996;32:347–364.
34. Finn B and Metcalfe J. Scaffolding feedback to maximize long-term error correction. *Memory & Cognition* 2010; **38:**951–961
35. Roach VA, Mistry MR, Wilson TD. Spatial visualization ability and laparoscopic skills in novice learners: evaluating stereoscopic versus monoscopic visualizations. *Anat Sci Educ.* 2014;7(4):295-301.
36. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial Abilities Training in Anatomy Education: A Systematic Review. *Anat Sci Educ.* 2020;13(1):71-79
37. 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.
38. 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.
39. 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.
40. 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.
41. 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.
42. Hegarty M and Waller DA. *Individual Differences in Spatial Abilities.* In P. Shah (Ed.) & A. Miyake, *The Cambridge Handbook of Visuospatial Thinking. 2005.* p. 121–169. Cambridge University Press

43. Hegarty M, Burte H, Boone AP. 2018. Individual differences in large-scale spatial abilities and strategies In: Montello DR (Editor). Handbook of Behavioral and Cognitive Geography. 1st Ed. Cheltenham, UK: Edward Elgar Publishing Limited; p 231–246.
44. Jordan K, Wüstenberg T, Heinze H-J, Peters M, Jäncke L. Women and men exhibit different cortical activation patterns during mental rotation tasks. Neuropsychologia 2002; 40:2397- 2408.
45. Peters M. Sex differences and the factor of time in solving mental rotation problems. Brain and Cognition 2005; 57(2):176–184
46. Kerkman DD, Wise JC and Harwood E. Impossible "mental rotation" problems a mismeasure of women's spatial abilities? Learning and Individual Differences 2001; 12(2000), 253–269.
47. Hirnstein M, Bayer U and Hausmann M. Sex-specific response strategies in mental rotation. Learning and Individual Differences 2009;19(2):225–228.
48. Alexander GM and Evardone M. Blocks and bodies: Sex differences in a novel version of the Mental Rotations Test. Hormones and Behavior 2008;53(1):177–184.
49. Campbell MJ, Toth AJ, Brady N. Illuminating sex differences in mental rotation using pupillometry. Bio Psychol 2018; 138:19-26
50. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia:: Lippincott Williams & Wilkins; 2008.
51. Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandenbroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. Clin Epidemiol. 2017;9:331-338.

# 9

# Summary
# Nederlandse Samenvatting
# Русское Резюме

# SUMMARY

In **chapter 1** a general introduction on three-dimensional visualization technology (3DVT) and its use in anatomical and surgical education is provided. Additionally, the role of visual-spatial abilities (VSA) in learning anatomy and surgical procedures is described. The overarching aim of this thesis was to gain evidence-based insights to improve anatomical and surgical education by evaluating how various levels of VSA interacts with learning using stereoscopic 3DVT.

In **chapter 2** we performed a systematic review and meta-analysis to compare the effectiveness of monoscopic 3D models with stereoscopic 3D models in teaching anatomy. We included 13 randomized controlled trials. Studies were grouped based on relative between-intervention differences in instructional methods and type of control conditions. When interactive, stereoscopic 3D models were compared to interactive monoscopic 3D models within a single level of instructional design by isolating stereopsis as the only true manipulated element in the experimental design, an effect size of 0.53 ($p < .00001$) was found. Stereoscopic vision had no effect in learning with non-interactive 3D images. With this comprehensive analysis we emphasized the importance of making a distinction between monoscopic and stereoscopic 3DVT.

Based on the evidence of the positive effect of stereoscopic vision, we performed a double-center randomized controlled trial to evaluate the effectiveness of stereoscopic vision in a 3D AR environment in teaching anatomy. In **chapter 3**, sixty (bio)medical students learned anatomy of the lower leg with either stereoscopic 3D AR model, monoscopic 3D desktop model or 2D images from anatomical atlas. As result, an aptitude-treatment interaction caused by VSA, as measured by the Mental Rotation Test, was observed. Students with low VSA achieved significantly higher posttest scores in the stereoscopic 3D AR group (49.2%) as compared to the monoscopic 3D desktop group (33.4%) and similar to the scores in the 2D group (46.4%). Students with high VSA performed equally well in all conditions. These differences were possibly attributed to the absence of stereoscopic vision in the monoscopic 3D desktop group. The findings indicated that VSA is able to compensate for poor instructional method as explained within the cognitive load theory. Since comparisons were made within various levels of instructional design, the true effect of stereoscopic vision in 3D AR environment still needed to be confirmed.

In **chapter 4**, we performed a follow-up study to evaluate the true effect of stereoscopic vision in a 3D AR environment by isolating it as the only true manipulated element between groups. In a double-center randomized controlled trial, sixty (bio)medical students learned anatomy of the lower leg with either stereoscopic 3D AR model or monoscopic 3D AR model. In the monoscopic condition, monoscopic view was obtained technically by projecting identical image to both left and right eyes. Both groups were blinded to the type of condition. As result, no significant differences were found between groups in terms of written knowledge (47.9 vs 49.1; p = .63) and specimen test scores (43.0 vs 46.3; p = .429), and perceived cognitive load scores (6.2 vs 6.2; p = .992). Regardless of intervention, VSA were positively associated with the specimen test scores ($\eta^2 = 0.13$, p = .003). These findings strongly suggested that stereoscopic vision was not the only one depth cue that have affected learning. Motion parallax (being able to walk around the model), as another important depth cue, could have compensated for the absence of stereoscopic vision in the control group.

The effect of stereoscopic vision was further evaluated in surgical training. In **chapter 5** we evaluated the effectiveness of watching an instructional video of a spatially complex procedure in 3D. In a randomized controlled trial, 108 surgical residents performed a five-flap Z-plasty on a simulation model after watching the instructional video either in 2D (monoscopically) or 3D (stereoscopically with active 3D glasses). As result, an aptitude-treatment interaction was observed. Overall, both groups performed equally well in terms of perceived cognitive load scores, performance scores and achieved safe lengthening. However, when accounted for VSA, only residents with high VSA benefitted from 3D visualization and achieved safe lengthening significantly more often than low VSA residents (OR = 6.6, p = .027). These findings indicate that VSA is able to interact with instructional method and enhance learning, as explained within the cognitive load theory.

The demonstrated aptitude-treatment interaction caused by VSA in learning with 3DVT (**chapter 3 and 5**) brought to light the importance of individualized approach in medical training. **In chapter 6**, we performed a randomized controlled trial and compared two types of intraoperative feedback (task-specific stepwise versus global rating scale) on performing a spatially complex procedure in fifty medical undergraduates in relation to VSA. As result, the task-specific stepwise feedback group performed significantly better than the global rating feedback group in terms of time in seconds (Δ 371 vs Δ 274; p = .027) and path length in meters (Δ 53.5 vs Δ 34.7; p = .046). However, when results were stratified by VSA, the greater improvement in time (p = .032) and path length (p = .053) was observed only in students with low VSA. Again, these findings demonstrated the aptitude-treatment interaction caused by VSA. More importantly, the findings demonstrated that alignment between feedback and instructional and learning activities improved learning, especially in students with low VSA.

Lastly, in **chapter 7**, we evaluated whether VSA can be improved by repeated practice of anatomy, starting from the early stages of medical training. In a case-control study, VSA of the first and second-year medical students (n = 45) was assessed before and after participation in a dissection course of ten weeks. The improvement in VSA scores were compared to students who did not participate in the course (n = 65). After ten weeks, both course participants and controls improved in their VSA scores. However, the improvement was significantly greater in students who participated in the dissection course (*first-year:* Cohen's $d$ = 0.41; *second-year:* Cohen's $d$ = 0.11). Additionally, the greatest improvement was observed in students with low VSA (Cohen's $d$ = 0.61). The findings indicate that VSA can be improved by repeated practice of anatomy, especially in individuals with low levels of VSA.

In **chapter 8,** the results of this thesis are put into a broader perspective and suggestions for future directions are made. The emphasis has been put on recognizing and accounting for the aptitude-treatment interaction caused by VSA in learning with stereoscopic 3DVT and its ability to be improved by repeated practice.

# NEDERLANDSE SAMENVATTING

**Hoofdstuk 1** vormt een algemene introductie en beschrijft de rol van drie-dimensionele (3D) technologie in anatomisch en chirurgisch onderwijs, het verschil tussen monoscopische en stereoscopische 3D visualisatie en hoe ruimtelijk inzicht zich verhoudt tot het bestuderen van anatomie en chirurgische procedures met 3D technieken. Het overkoepelende doel van dit promotieonderzoek was om nieuwe inzichten te verkrijgen in 3D leren in relatie tot ruimtelijk inzicht, om hiermee het anatomisch en chirurgisch onderwijs te verbeteren.

**Hoofdstuk 2** bevat een samenvatting en meta-analyse van de literatuur betreffende het leereffect van stereoscopische 3D visualisatie in anatomisch onderwijs. Er werden in totaal 13 onderzoeken in het systemische review geïncludeerd. De bestudeerde uitkomstmaat was de gemiddelde score op de schriftelijke anatomische kennistoets. Dit review laat zien dat stereoscopische 3D visualisatie effectief is, maar alleen wanneer deze wordt toegepast binnen de interactieve 3D leeromgeving: interactieve stereoscopische 3D modellen zijn effectiever dan interactieve monoscopische 3D modellen (effectgrootte van 0.53; $p < .0001$) bij het leren van de anatomie. Het werkingsmechanisme kan mogelijk worden verklaard vanuit de cognitieve belasting theorie: monoscopische visualisatie van 3D anatomie leidt tot hogere mentale belasting bij studenten met lager ruimtelijk inzicht, met een negatief effect op het leerproces als gevolg.

**Hoofdstuk 3** beschrijft een experiment onder zestig (bio)medische studenten waarin de effectiviteit van stereoscopische visualisatie in een interactieve 3D augmented reality (AR) omgeving wordt onderzocht in relatie tot ruimtelijk inzicht. Drie verschillende leermethoden werden met elkaar vergeleken: stereoscopische 3D AR visualisatie van het onderbeen, monoscopische 3D visualisatie op een desktop van het onderbeen en 2D visualisatie via afbeeldingen uit een anatomische atlas. De opgedane anatomische kennis werd getoetst door middel van een gevalideerde schriftelijke anatomie kennistoets. De resultaten laten zien dat ruimtelijk inzicht invloed heeft op het leren van de anatomie en dat het een zogenaamd ‘aptitude-treatment effect’ veroorzaakt: alleen studenten met een beperkt ruimtelijk inzicht hebben baat bij het leren met het stereoscopische 3D AR model ten opzichte van het monoscopische 3D desktopmodel (49% versus 33.4%). Studenten met goed ruimtelijk inzicht leerden even goed met alle drie de leermethoden. Deze verschillen zijn mogelijk toe te schrijven aan de afwezige stereopsis, oftewel de diepteperceptie, binnen de monoscopische 3D desktop model groep.

**Hoofdstuk 4** beschrijft een vervolgexperiment onder zestig (bio) medische studenten waarin wij het daadwerkelijke effect van stereoscopische visualisatie in een 3D AR omgeving hebben onderzocht in relatie tot ruimtelijk inzicht. Deelnemers bestudeerden de anatomie van het onderbeen door middel van een monoscopisch 3D model of een stereoscopisch 3D model binnen dezelfde 3D AR omgeving. De opgedane anatomische kennis werd getoetst met een schriftelijke kennistoets en een praktijktoets op kadavers. Daarnaast werd ook de mate van mentale belasting gemeten door middel van een gevalideerde (NASA-TLX) vragenlijst. De uitkomsten laten zien dat de monoscopische weergave, niet leidt tot een slechtere leerprestatie (schriftelijke toets: 47.9 vs 49.1; p = .635; praktijktoets: 43.0 vs 46.3; p = .429) of hogere mentale belasting (6.2 vs 6.2; p = .992). De bewegingsparallax, oftewel het zich kunnen verplaatsen in de ruimte rondom het model, heeft mogelijk bijgedragen aan het compenseren van de uitgeschakelde diepteperceptie.

In **hoofdstuk 5** hebben wij het effect van stereoscopische visualisatie op het leren van chirurgische procedures onderzocht in relatie tot ruimtelijk inzicht. 108 arts-assistenten chirurgie werden verdeeld over twee groepen: groep 1 bekeek de instructievideo van een ruimtelijk complexe procedure in 2D (monoscopisch), groep 2 bekeek de identieke video in 3D (stereoscopisch, met behulp van gepolariseerde 3D brillen). Na het bekijken van de instructievideo hebben alle deelnemers de procedure uitgevoerd op een simulatiemodel. Hun prestatie werd gemeten en uitgedrukt in de gemiddelde score op de gestandaardiseerde OCHRA-checklist en het gemiddelde percentage van veilige uitvoering. Ook werd de mentale belasting tijdens bestuderen en uitvoeren van de procedure gemeten door middel van een gevalideerde (NASA-TLX) vragenlijst. De resultaten tonen aan dat ruimtelijk inzicht van invloed is op het leren van chirurgische procedure met behulp van 3D technologie: het bekijken van de instructievideo in 3D bleek alleen effectief te zijn voor arts-assistenten met een hoog ruimtelijk inzicht met betrekking tot veilige uitvoering van de procedure (odds ratio = 6.67; p = .027).

**Hoofdstuk 6** beschrijft een experiment waarin het leereffect van intra-operatieve feedback op het uitvoeren van een open liesbreuk procedure op een simulatiemodel onderzocht werd in relatie tot ruimtelijk inzicht. Hierbij hebben wij onderscheid gemaakt tussen twee typen feedback: stapsgewijze taak-specifieke (OCHRA) feedback versus globale (OSATS) feedback. Uit dit onderzoek is gebleken dat OCHRA-feedback effectiever is voor studenten met laag ruimtelijk inzicht wanneer het gaat om het verbeteren van de snelheid in seconden waarmee zij de procedure uitvoeren (Δ371 vs Δ274; p = .027) en de efficiëntie van de handbewegingen uitgedrukt in meters (Δ53.5 vs Δ34.7; p = .046). Dit benadrukt het belang van 'constructive alignment', waarbij de manier waarop feedback wordt gegeven nauw aansluit op de manier waarop de leerdoelen en activiteiten zijn opgesteld en uitgevoerd. Ruimtelijk inzicht lijkt een belangrijke rol te spelen in het leren van anatomie en chirurgische procedures met stereoscopische 3D visualisatie.

In **hoofdstuk 7** hebben wij onderzocht of ruimtelijk inzicht verbeterd kan worden door repetitieve deelname aan het anatomisch onderwijs. Het ruimtelijk inzicht van de eerste- en tweedejaars medische studenten werd gemeten vóór en na hun deelname aan een tien-weekse dissectiecursus door middel van een gevalideerde Mentale Rotatie Test. Hun scores op de ruimtelijk inzicht test werden vervolgens vergeleken met de scores van hun medestudenten die niet deelnamen aan deze cursus, waarbij cursisten een betere score bleken te hebben (*1$^{e}$-jaars*: Δ6.0 vs Δ4.1, $p = .019$; *2$^{e}$-jaars*: Δ6.5 vs Δ6.1; $p = .03$). Met name studenten met laag ruimtelijk inzicht lieten een significatie verbetering zien in de scores (Δ8.4 vs Δ6.8, $p = .001$, Cohen's $d = 0.61$). Resultaten van deze studie tonen aan dat ruimtelijk inzicht verbeterd kan worden door actieve en repetitieve deelname aan anatomisch onderwijs.

In **hoofdstuk 8** worden de resultaten van dit proefschrift in een breder perspectief geplaatst en bediscussieerd. Ook worden de aanbevelingen voor anatomisch en chirurgisch onderwijs gedaan, waarbij de nadruk wordt gelegd op het herkennen van de relatie tussen het individuele niveau van ruimtelijk inzicht en het type 3D visualisatie bij het leren van anatomie en chirurgische procedures.

# РУССКОЕ РЕЗЮМЕ

В **главе 1** описывается общее введение в технологию трехмерной визуализации, разница между монокулярным и стереоизображением и ее использование в анатомическом и хирургическом образовании. Дополнительно описывается роль зрительно-пространственных способностей (ЗПС) в изучении анатомии и хирургических процедур. Основная цель этой диссертации, озаглавленной «Единого стандартного подхода не существует», состояла в том, чтобы получить основанные на доказательствах идеи для развития анатомического и хирургического образования. Это было достигнуто путем оценки того, как различные уровни ЗПС взаимодействуют в обучении с использованием стереоизображения в трёхмерных технологиях, которые создают иллюзию трехмерной глубины из заданных двухмерных изображений.

В **главе 2** представлен систематический обзор и мета-анализ исследований, в которых сравнивается обучающий эффект монокулярного и стереоизображения в рамках трехмерной технологий. Результаты нашего анализа показывают, что стереоизображение интерактивных анатомических трехмерных моделей более эффективно для изучения анатомии, чем монокулярное изображение (размер эффекта 0,53). Эти различия объясняются в рамках теории когнитивной нагрузки, в которой выделяются два разные когнитивные процессы. Основываясь на этих выводах, мы подчеркнули важность стереоизображения в трёхмерных технологиях.

Основываясь на доказательствах положительного эффекта стереоизображения, мы провели двухцентровое рандомизированное контролируемое исследование **(глава 3).** В этом эксперименте мы исследовали эффективность стереоизображения в рамках дополненной реальности (ДР). Мы сравнили три метода обучения: трехмерную ДР модель со стереоизображением, трехмерную компьютерную модель без стереоизображения и двухмерные изображения из анатомического атласа. В этом исследовании приняли участие шестьдесят студентов-медиков. Результаты показывают, что ЗПС влияет на обучение и вызывает так называемое «взаимодействие способностей и лечения»: только учащиеся с низким уровнем ЗПС извлекли пользу из обучения со трехмерной ДР моделью по сравнению с трехмерной компьютерной моделью (49% против 33%). Учащиеся с высоким уровнем ЗПС извлекли пользу из всех методов обучения. По всей вероятности, эти различия были связаны с отсутствием стереоизображения в группе трехмерной компьютерной модели. Поскольку сравнения проводились на разных уровнях учебного дизайна, истинный эффект стереоизображения в рамках ДР все еще подлежал подтверждению.

В **главе 4** мы провели дополнительное исследование, чтобы оценить истинный эффект стереоизображения в рамках ДР. В ходе двухцентрового рандомизированного контролируемого исследования, шестьдесят студентов-медиков были разделены на две группы: одна группа изучала анатомию голени с помощью трехмерной ДР модели со стереоизображением, другая группа изучала анатомию с помощью той же трехмерной ДР -модели, но с монокулярном изображением. В контрольном режиме монокулярное изображение было технически получено путем проецирования идентичного изображения на левый и правый глаз. Студенты обеих групп не были проинформированы и не знали о типе изображения. В результате не было обнаружено значительных различий между группами с точки зрения приобретённых знаний и воспринимаемой когнитивной нагрузки. Независимо от типа обучения, ЗПС было положительно связано с приобретёнными знаниями по анатомии голени. Эти результаты убедительно свидетельствуют о том, что стереоизображение следовало не единственным источником информации о трехмерной глубине, которая могла повлиять на обучение. Монокулярный параллакс движения (возможность ходить вокруг трёхмерной модели в физическом пространстве), как еще один важный источник информации о трехмерной глубине, смог наверняка компенсировать отсутствию стереоизображения в контрольной режиме.

В **главе 5**, влияние стереоизображения было дополнительно исследовано для обучения хирургических процедур с помощью обучающих видео. 108 аспирантов хирургии были разделены на две группы: одна группа смотрела пространственно обучающее видео сложной процедуры в двумерном (монокулярном) изображении, другая группа смотрела то же самое видео в трёхмерном изображении (стереоизображении). После просмотра обучающего видео, все участники выполнили процедуру на имитационной модели. Наши результаты снова показали, что ЗПС влияет на обучение с использованием стереоизображения и вызывает вышеупомянутое взаимодействие: просмотр обучающего видео в трёхмерном изображении оказался эффективным только для аспирантов с высоким уровнем ЗПС. Описанное взаимодействие между ЗПС и методом обучения объясняется в рамках теории когнитивной нагрузки.

Продемонстрированное взаимодействие между ЗПС и методом обучения **(глава 3 и 5)** выявило важность индивидуального подхода в медицинском обучении. В **главе 6** описывается эксперимент, в котором мы исследовали влияние интраоперационной обратной связи на выполнение пространственно сложной хирургической процедуры в отношении ЗПС. Мы различили два метода обратной связи: поэтапную и ориентированную на конкретную задачу и глобальную обратную связь. В этом исследовании приняли участие пятьдесят студентов-медиков, которые дважды выполнили операцию по удалению паховой грыжи на имитационной модели. Обратная связь была предоставлена между двумя процедурами. В результате группа поэтапной обратной связи выполняла операцию значительно лучше, чем группа глобальной обратной связи с точки зрения скорости и радиуса движения рук. Однако, когда результаты были стратифицированы в соответствии с уровнем ЗПС, улучшение наблюдалось только у студентов с низким уровнем ЗПС. Опять же, эти результаты продемонстрировали взаимодействие между ЗПС и методом обучения. Дополнительно, результаты подчеркивают важность принципа «конструктивного согласования», согласно которому метод обратной связи должен быть согласован с методом формулирования и реализации целей учебного материала.

Таким образом, ЗПС играет важную роль в обучении анатомии и хирургических процедур с использованием технологии трехмерной визуализации. Фактически, ЗПС взаимодействует с методом обучения и, таким образом, влияет на индивидуальные результаты обучения.

Наконец, в **главе 7** мы исследовали возможность развития ЗПС путем интенсивного обучения анатомии, начиная с ранних лет медицинского обучения. В рамках этого исследования мы измерили уровень ЗПС студентов-медиков первого и второго курса до и после их участия в 10-недельном курсе анатомического вскрытия. Затем мы сравнили их результаты по тесту на ЗПС с результатами их сокурсников, которые не участвовали в этом курсе. По окончании курса студенты набрали больше баллов по тесту, чем их сокурсники (размер эффекта 0,41 (первый курс), 0,11 (второй курс)). Особенно значительно улучшились результаты по тесту у студентов с низким уровнем ЗПС (размер эффекта 0.61). Результаты этого исследования показывают, что ЗПС склонно к развитию путем повторной практикой анатомии.

В **главе 8** результаты этой диссертации рассматриваются и обсуждаются в более широкой перспективе. Даны также рекомендации по анатомическому и хирургическому образованию. Акцент делается на распознавании и учете взаимодействия между ЗПС и обучением с использованием стереоизображения в трёхмерных технологиях, развитие ЗПС, а так же индивидуальный подход в медицинском обучении.

S

# Supplementary

# SUPPLEMENTARY

**Supplementary material 1.** Characteristics of included studies.

| Author | Study design | Participants | N | SV assessed | VSA assessed | Anatomical region |
|---|---|---|---|---|---|---|
| Al-Khalili et al., 2014[45] | RCT pre-posttest | Veterinary students | 84 | no | no | Thorax, abdomen, pelvis (canine) |
| Bogomolova et al., 2019[3] | RCT posttest only | Medical and biomedical students | 58 | No | Yes, by MRT | Lower limb |
| Cui et al., 2016[4] | RCT pre-posttest | Medical students | 39 | no | yes, by MRT | Head and neck vascular anatomy |
| de Faria et al., 2016[40] | RCT pre-posttest | Graduate medical students | 84 | no | no | Neuroanatomy |
| Ekstrand et al., 2018[43] | RCT pre-posttest | Medical students | 64 | no | no | Neuroanatomy |
| Goodarzi et al., 2017[34] | non RCT pre-posttest | Students of School of Education | 249 | no | no | Skull |
| Hackett and Proctor 2018[41] | RCT pre-posttest | Nursing students | 179 | no | no | Cardiac anatomy |
| Hilbelink 2009[35] | non RCT pre-posttest | Nursing students, wellness program students | 248 | no | no | Skull |
| Kockro et al., 2015[46] | RCT posttest only | Medical students | 169 | no | no | Ventricular system |

| Intervention, technology | Feature intervention | Control(s), technology | Feature control | Cognitive level anatomy questions |
|---|---|---|---|---|
| Stereoscopic 3D dissection images on video, desktop with anaglyphic 3D glasses | Non-interactive | 1) 2D dissection images on video, desktop<br>2) 2D images, textbook | Non-interactive | High order |
| 1) Stereoscopic 3D AR model, Hololens | Interactive user control | 1) Monoscopic 3D model, desktop<br>2) 2D anatomical atlas | 1) interactive user control<br>2) non-interactive | combination |
| Stereoscopic 3D model, stereo-projector screen with polarizing 3D glasses | Interactive, instructor control | 2D images, projector screen | Non-interactive | Combination |
| Stereoscopic 3D model on videoclip, desktop with stereoscopic anaglyphic 3D glasses | Interactive user control | 1) Monoscopic 3D model on videoclip, desktop<br>2) 2D images, projector screen | 1) Interactive user control<br>2) non-interactive | Low order |
| Stereoscopic 3D model, VR with Vive HTC head-mounted display | Interactive user control | 2D images, paper-based booklet | Non-interactive | High order |
| Stereoscopic 3D images on video, unknown medium with anaglyphic 3D glasses | Non-interactive | 2D images on video, unknown medium | Non-interactive | Unknown |
| Stereoscopic 3D model, photopolymer | Interactive user control | 1) Monoscopic 3D model, desktop<br>2) 2D images, paper based | 1) Interactive user control<br>2) Non-interactive | Low order |
| Stereoscopic 3D images, desktop with anaglyphic 3D glasses | Non-interactive | 2D images, desktop | Non-interactive | Low order, High order |
| Stereoscopic 3D images, stereo-projector screen with polarizing glasses | Non-interactive | 2D images, projector screen | Non-interactive | High order |

**Supplementary material 1.** Continued.

| Author | Study design | Participants | N | SV assessed | VSA assessed | Anatomical region |
|---|---|---|---|---|---|---|
| Luursema et al., 2017[37] | RCT pre-posttest | Medical and biomedical students | 63 | yes | Yes, by MRT | Neck anatomy |
| Luursema et al., 2006[38] | RCT posttest only | Students and employees of behavioral sciences. | 36 | yes | Yes, by MRT | Abdomen |
| Luursema et al., 2008[2] | RCT posttest only | Students and employees of behavioral sciences. | 46 | yes | Yes, by MRT | Abdomen |
| Moro et al., 2017[36] | RCT posttest only | Medical, biomedical and health sciences | 59 | no | No | Skull |
| Remmele et al., 2018[39] | RCT cross-over posttest only | Teacher trainees of biological and non-biological education | 171 | yes | No | Skull and ear |
| Stepan et al., 2017[44] | RCT pre-posttest | Medical students | 66 | no | no | Neuroanatomy |

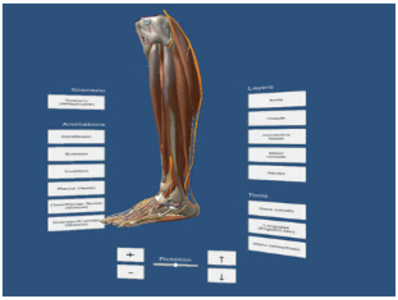
| Intervention, technology | Feature intervention | Control(s), technology | Feature control | Cognitive level anatomy questions |
|---|---|---|---|---|
| Stereoscopic 3D model, VR with Oculus Rift SDK 2 head-mounted display | Interactive user control | 1) Monoscopic 3D model, VR with Oculus Rift head-mounted display<br>2) Watching virtual sea world, Oculus Rift head-mounted display | Interactive user control | High order |
| Stereoscopic 3D model, desktop with 3D shutter glasses | Interactive user control | 2D images, desktop | Non-interactive | High order |
| Stereoscopic 3D model, desktop with 3D shutter glasses | Interactive user control | Monoscopic 3D model, desktop | Interactive user control | High order |
| Stereoscopic 3D model, VR with Oculus Rift head-mounted display | Interactive user control | 1) Monoscopic 3D model, AR with Samsung Galaxy Tab S2 tablet<br>2) Monoscopic 3D model, Samsung Galaxy Tab S2 tablet | 1) Interactive user control<br>2) Interactive user control | Low order |
| 1) Dynamic stereoscopic 3D model, desktop with 3D shutter glasses<br>2) Static stereoscopic 3D model, desktop with 3D shutter glasses | 1) Interactive user control<br>2) non-interactive user control | 1) Dynamic monoscopic 3D model, desktop<br>2) Static monoscopic 3D model, desktop | 1) Interactive user control<br>2) non-interactive user control | High order |
| Stereoscopic 3D model, VR with Oculus Rift head-mounted display | Interactive user control/ non-interactive user control | 2D images, desktop | Non-interactive | Low order |

**Supplementary material 1.** Continued.

| Author | Study design | Participants | N | SV assessed | VSA assessed | Anatomical region |
|---|---|---|---|---|---|---|
| Wainman et al., 2019[42] | RCT Posttest only | Health science, Arts and Science, Engineering, Life Science, Science, Social Science, Humanities, Computer Science students | 80 | No | Yes, by MRT | Pelvis |

N = number of participants; SV = stereovision, VSA = visual-spatial abilities, MRT = Mental Rotation Test; RCT = randomized controlled trial; 3D = three-dimensional; 2D = two-dimensional; VR = virtual reality; AR = augmented reality; MR = mixed reality

**Supplementary material 2.** *DynamicAnatomy* application for Hololens®: a stereoscopic 3D holographic model of the lower leg.



**Supplementary material 3.** *DynamicAnatomy* application for Hololens®: a *DynamicAnatomy* application for Windows: a monoscopic 3D desktop model of the lower leg.

| Intervention, technology | Feature intervention | Control(s), technology | Feature control | Cognitive level anatomy questions |
|---|---|---|---|---|
| 1) Stereoscopic 3D model, VR with HTC Vive<br>2) Stereoscopic 3D model, MR with Hololens | Interactive user control | 1) Monoscopic 3D model, VR with HTC Vive<br>2) Monoscopic 3D model, MR with Hololens | Interactive user control | Low order |

**Supplementary material 4A.** Learning objectives.

At the end of the learning session, students should be able to:

1. Identify the following bones of the lower leg, ankle, and foot:

| | | |
|---|---|---|
| tibia | navicular | lateral cuneiform |
| fibula | cuboid | metatarsals 1-5 |
| talus | medial cuneiform | phalanges 1-5 |
| calcaneus | intermediate cuneiform | |

2. Identify the following muscles including their origins and insertions:

| **I** | **II** | **III** | **IV** |
|---|---|---|---|
| m. gastrocnemius (medial head) | m. tibialis posterior | m. tibialis anterior | m. peroneus longus |
| m. gastrocnemius (lateral head) | m. flexor digitorum longus | m. extensor hallucis longus | m. peroneus brevis |
| m. soleus | m. Flexor hallucis longus | m. extensor digitorum longus | |
| m. plantaris | | m. peroneus tertius | |

3. Indicate for each of the two ankle joints (tibiotalar and subtalar/talocalcaneal) which of the following movements they facilitate

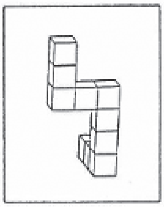
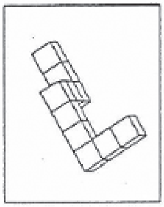
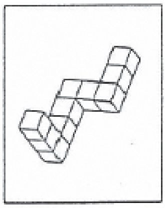
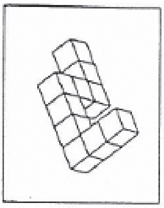
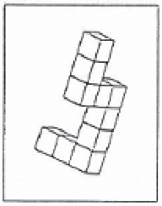
   dorsiflexion = flexion(to bend) of the foot
   plantar flexion = extension(to stretch) of the foot
   inversion= inner rotation of the foot
   eversion= external rotation of the foot

4. Determine the course of the following muscles in relation to neighboring muscles and bones
    - m. soleus
    - m. tibialis anterior
    - m. tibialis posterior
    - m. peroneus longus
5. Indicate for each muscle which movements (dorsiflexion, plantar flexion, inversion, eversion) they facilitate. This is determined by their origins and insertions, and by their course with regard to the joint axes.
6. Determine which structure (joint, bone and/or muscle) is injured in case of a patient with affected range of movement in the ankle.

**Supplementary material 4B.** Instructions for learning session.

1. Identify the relevant bones
2. Identify the two joints of the ankle: the *tibiotalar* (between tibia and talus) and *subtalar or talocalcaneal (*between talus and calcaneus) joints. Note that each of the two ankle joints have their own unique joint axes, and they determine the motion patterns together with the origin, insertion, and course of the muscles
3. Relate the joint axes to the 4 movement patterns.
   There are 14 muscles involved in the movement of the ankle. They are organized in four compartments (see table): I = Superficial posterior (= back side), II = Deep posterior, III = Anterior (= front side), IV = Lateral (=outer side).
4. Identify the m. tibialis anterior and its direct neighboring bones and muscles
5. Examine its course: from the upper head of the tibia (its origin) to the lower side of the medial cuneiform of the foot (its insertion), through the medial side (= inner) side of the ankle joints. Therefore, contraction of this muscle results in inversion (= inner rotation of the foot), and in plantar flexion (= extension of the foot) due to its insertion on the side of the foot pad.
6. Repeat steps 4 and 5 for the other muscles. Note that their unique course and insertion determine which movement they facilitate

**Supplementary material 5A.** Explanation of the Mental Rotation Test.
Look at these five figures

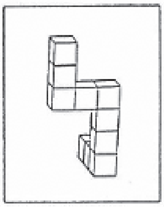






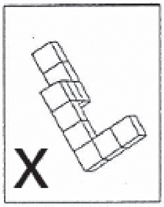
Note that these are all pictures of the same object which is shown from different angles. Try to imagine moving the object, as you look from one drawing to the next.

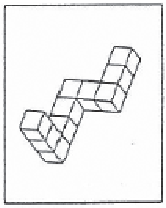
In the following 24 questions you will be asked to identify the two drawings which show the same object.

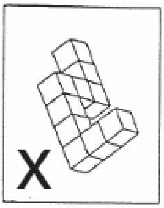
EXAMPLE:

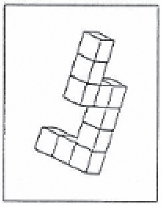
Now look at this subject. Two of these four drawings show the same object. They are marked with X:











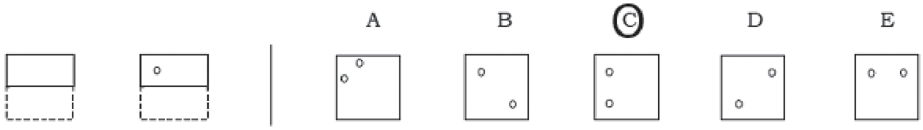
**Supplementary material 5B.** Explanation of the Paper Folding Test.

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line.

The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point.

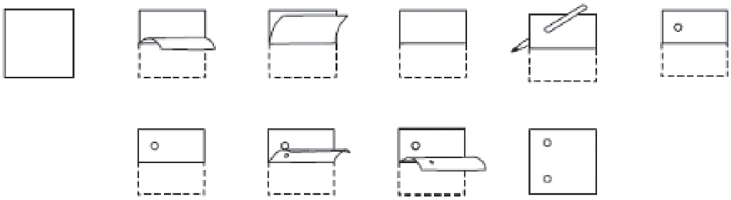
One of the five figures on the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and encircle the correct option.

EXAMPLE: (In this problem only one hole was punched in the folded paper).



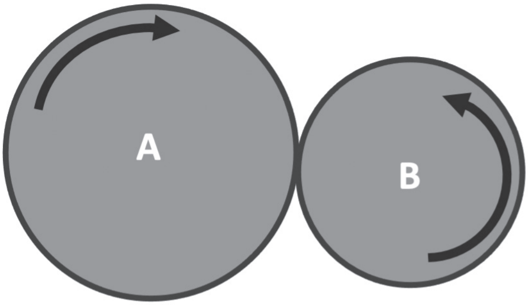
The correct answer to the sample problem above is **C**.

The figures below show how the paper was folded and why **C** is the correct answer.

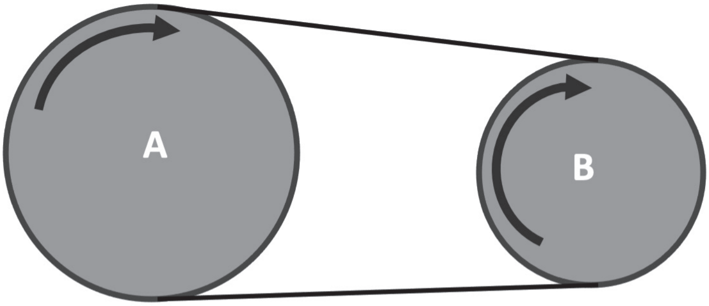


**Supplementary material 5C.** Explanation of the Mechanical Reasoning Test.
In this test you will be presented with gear problems. A gear problem consists of 2 or more connected gears, which can rotate in response to one another.

In the example below, you can see how gear A rotates in a clockwise direction. As a result, gear B will rotate in a counter clockwise direction.



Instead of being connected directly, these gears can also be connected by a wire, as shown in the example below. Here gear B will rotate in the same direction as gear A.



On the next pages, you will be presented with items concerning these gear problems. For each item, the question is: **Are the arrows in the figure in the correct direction or not?** The first and last gear of the problem have been provided with an arrow, which are in the correct or incorrect position.

There are 12 problems in total, you have 3 minutes to complete as many items as you can. Focus on both accuracy and speed of your responses.

**Supplementary material 6.** Anatomical knowledge test: an example of extended matching questions and open-ended questions in the factual, functional and spatial domains.

**Factual knowledge domain**

Which muscle is indicated?

M. extensor digitorum longus
M. extensor hallucis longus
M. flexor digitorum longus
M. flexor hallucis longus
M. gastrocnemius lateral head
M. gastrocnemius medial head
M. peroneus brevis
M. peroneus longus
M. peroneus tertius
M. plantaris
M. soleus
M. tibialis anterior
M. tibialis posterior

**Functional knowledge domain**

Jan is playing beach volleyball. While jumping off the ground, he hears a snapping sound followed by an immediate sharp pain on the back side of his lower leg. He is not able to walk on his toes anymore.

Which muscle is injured and why?

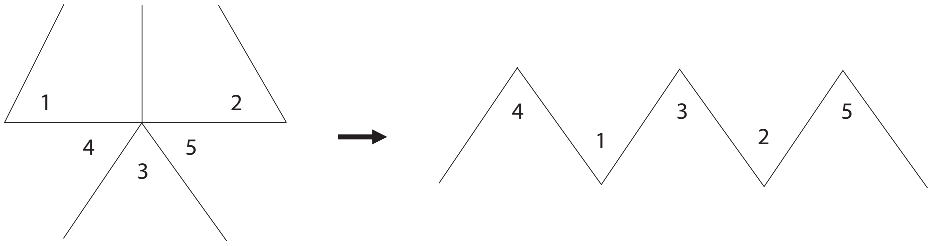
**Spatial knowledge domain**

Which 4 bones, that are listed below, are in direct contact with talus?

Tibia
Fibula
Calcaneus
Cuboid
Intermediate cuneiform
Lateral cuneiform
Medial Cuneiform
Navicular

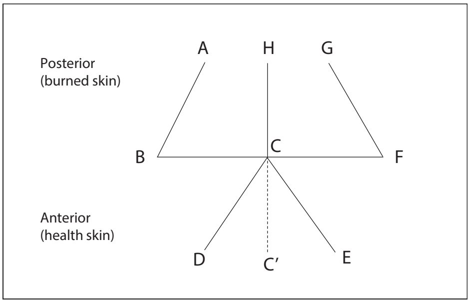
**Supplementary material 7.** Anatomical specimen test: an example of open ended questions in spatial and functional knowledge domains.

| STATION 1 | |
|---|---|
| Question | Answer |
| What is the name of **structure 1**? | |
| What is the name of **structure 2**? | |
| What is the name of **structure 3**? | |
| What is the name of **structure 4**? | |
| STATION 2 | |
| Question | Anwser |
| Which movement of the **ankle** joint is facilitated by **structure 11**? | |
| Which movement of the **ankle** joint is facilitated by **structure 12**? | |
| Next to *dorsiflexion*, which movement of the **ankle** joint is facilitated by **structure 13**? | |
| Next to *plantar flexion*, which movement of the **ankle** joint is facilitated by **structure 14**? | |



**Supplementary material 8.** The 5-flap Z-plasty.

**Supplementary material 9.** OCHRA checklist for the assessment of a 5-flap Z-plasty.



Total number of steps: 10. Total score: 10.

| | **Surgical steps** | | | | **Executional error** | |
|---|---|---|---|---|---|---|
| | **Step** | **Substep (structure)** | **Action** | | Correctly performed = 0.5, 1, or 1.5<br>Incorrectly performed = 0 | |
| **OCHRA 1** | 1. Flap design | A. Skin | 1. Identify | Identify resting skin tension lines, the lines of maximal extension, and the slack to determine how and where the jumping man plasty will be positioned. | DCE flap position on the burn site of the skin (posterior) | **1** |
| **OCHRA 2** | | | | | CC' = CH = 2.5 cm | **1** |
| **OCHRA 3** | | | 2. Mark | Mark the jumping man plasty that consists of two opposing Z-plasties with a YV-plasty in between. | DCE angle exceeds 60–90˚ | **1** |
| **OCHRA 4** | | | | | GFC angle exceeds 45–60˚ | **1** |
| **OCHRA 5** | | | | | ABC angle exceeds 45–60˚ | **1** |

**Supplementary material 9.** OCHRA checklist for the assessment of a 5-flap Z-plasty.

| | Surgical steps | | | | Executional error | |
|---|---|---|---|---|---|---|
| | **Step** | **Substep (structure)** | **Action** | | Correctly performed = 0.5, 1, or 1.5<br>Incorrectly performed = 0 | |
| **OCHRA 6** | 2. Flap creation | A. Skin | 1. Incise | Incise the skin over the marked lines. | The skin is not completely incised and/or not incised over the marked lines | **0.5** |
| **OCHRA 7** | | | 2. Transpose | Transpose both the Z-plasties by interchanging the location of the created triangles per Z-plasty | Transposition of Z-plasties in a wrong direction | **1.5** |
| **OCHRA 8** | | | 3. Trim | Adjust the size of the Z-plasty flaps | Trimming of the flaps of the healthy skin | **0.5** |
| **OCHRA 9** | | | 4. Advance | Advance the YV-plasty. | Incorrect advancement | **1.5** |
| **OCHRA 10** | 3. Wound closure | A. Skin | 1. Close | Close the skin with standing transcutaneous sutures. First, the three-point sutures should be placed. | Not all five wound sites are closed | **1** |
| **Total** | | | | | | **10** |

**Supplementary material 10.** The logistic regression model and coefficients for the outcome measure "safe lengthening".

| | | | 95% CI for Exp (B) | |
|---|---|---|---|---|
| | Sig | Exp (B) | Lower | Upper |
| Intervention (stereoscopic vs. monoscopic) | 0.078 | 0.337 | 0.100 | 1.131 |
| VSA (high VSA vs. low VSA) | 0.011 | 0.219 | 0.068 | 0.704 |
| VSA x Intervention | 0.027 | 6.671 | 1.239 | 35.904 |

Exp (B)=odds ratio

Calculated odds ratios based on the values from the model:

- 3D & high VSA = 8.8
- 3D & Low VSA = 1.8
- 2D & high VSA = 1.9
- 2D & low VSA = 1.6

**Supplementary material 11.** OCHRA checklist for open inguinal hernia repair.

**Surgical steps**

| **Step** | **Substep** | **Action** |
|---|---|---|
| *1. External oblique aponeurosis exposure* | A. Skin | 1. Incise |
| | B. Subcutaneous tissue | 1. Incise |
| | C. Superficial epigastric vein | 1. Transect |
| | D. Scarpa's fascia | 1. Incise |
| | E. Subcutaenous tissue | |
| *2. Inguinal canal exposure* | A. External oblique aponeurosis | 1. Identify |
| | | 2. Incise |
| | | 3. Dissect |
| *3. Spermatic cord mobilization* | A. Spermatic cord | 1. Isolate |
| | | 2. Encircle |
| *4. Hernia sac resection* | A. Hernia sac | 1. Identify |
| | | 2. Remove |
| *5. Mesh placement* | A. Inguinal canal | 1. Expose |
| | B. Mesh | 1. Trim mesh |
| | | 2. Position mesh parallel to inguinal ligament |
| | | 3. Fixate - medial to rectus sheath |
| | | 4. Fixate - caudal to inguinal ligament |
| | | 5. Split - superior 2/3 inferior 1/3 |
| | | 6. Position - folding tails correctly |
| | | 7. Fixate - lateral |
| | | 8. Trim mesh laterally |
| | | 9. Position mesh under aponeurosis |
| | | 10. Fixate - cranial to internal oblique muscle |
| *6. Wound closure* | A. External oblique aponeurosis | 1. Close |
| | B. Scarpa's fascia | 1. Close |
| | C. Skin | 1. Close |

| Performed correctly? | Procedural error | Executional error | Consequential? |
|---|---|---|---|
| | | | |
| | | | HAZARD – Iliohypogastric nerve damage<br>HAZARD - Superficial epigastric vessels damage |
| | | | HAZARD - Ilioinguinal nerve damage |
| | | | HAZARD - Genital branch of genitofemoral nerve |
| | | | |
| | | | HAZARD - Pubic periosteum damage<br>HAZARD - Femoral vessels and nerve damage<br>HAZARD - prosthetic inguinal ring too wide or too small<br>HAZARD - Iliohypogastric nerve damage |
| | | | |

**Supplementary material 12**. OSATS checklist.

| | | |
|---|---|---|
| Respect for tissue | 1<br>Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments | 2 |
| Time and motion | 1<br>Many unnecessary moves | 2 |
| Instrument handling | 1<br>Repeatedly makes tentative or awkward moves with instruments | 2 |
| Knowledge of instruments | 1<br>Frequently asked for the wrong instrument or used an inappropriate instrument | 2 |
| Use of assistance | 1<br>Consistently placed assistants poorly or failed to use assistants | 2 |
| Flow of operation and forward planning | 1<br>Frequently stopped operating or needed to discuss next move | 2 |
| Knowledge of specific procedure | 1<br>Deficient knowledge. Needed specific instruction at most operative steps | 2 |

| | | |
|---|---|---|
| 3<br>Careful handling of tissue but occasionally caused inadvertent damage | 4 | 5<br>Consistently handled tissues appropriately with minimal damage |
| 3<br>Efficient time/motion but some unnecessary moves | 4 | 5<br>Economy of movement and maximum efficiency |
| 3<br>Competent use of instruments although occasionally appeared stiff or awkward | 4 | 5<br>Fluid moves with instruments and no awkwardness |
| 3<br>Knew the names of most instruments and used appropriate instrument for the task | 4 | 5<br>Obviously familiar with the instruments required and their names |
| 3<br>Good use of assistants most of the time | 4 | 5<br>Strategically used assistant to the best advantage at all times |
| 3<br>Demonstrated ability for forward planning with steady progression of operative procedure | 4 | 5<br>Obviously planned course of operation with effortless flow from one move to the next |
| 3<br>Knew all important aspects of the operation | 4 | 5<br>Demonstrated familiarity with all aspects of the operation |

# List of Publications

# LIST OF PUBLICATIONS

**Bogomolova K**, van Merriënboer JJG, Sluimers JE, Donkers J, Wiggers T, Hovius SER, van der Hage JA. The effect of a three-dimensional instructional video on performance of a spatially complex procedure in surgical residents in relation to their visual-spatial abilities. Am J Surg 2021;222(4):739-745.

Nazari T, **Bogomolova K**, Ridderbos M, Dankbaar MEW, van Merriënboer JJG, Lange JF, Wiggers T, van der Hage JA. Global versus task-specific postoperative feedback in surgical procedure learning. Surgery 2021;170(1):81-87.

**Bogomolova K**, Sam AH, Misky AT, Gupte CM, Strutton PH, Hurkxkens TJ, Hierck BP. Development of a Virtual Three-Dimensional Assessment Scenario for Anatomical Education. Anat Sci Educ 2021;14(3):385-393.

**Bogomolova K**, Hierck BP, Looijen AEM, Pilon JNM, Putter H, Wainman B, Hovius SER, van der Hage JA. Stereoscopic three-dimensional visualisation technology in anatomy learning: A meta-analysis. Med Educ 2021;55(3):317-327.

**Bogomolova K**, van der Ham IJM, Dankbaar MEW, van den Broek WW, Hovius SER, van der Hage JA, Hierck BP. The Effect of Stereoscopic Augmented Reality Visualization on Learning Anatomy and the Modifying Effect of Visual-Spatial Abilities: A Double-Center Randomized Controlled Trial. Anat Sci Educ 2020;13(5):558-567.

**Bogomolova K**, Hierck BP, van der Hage JA, Hovius SER. Anatomy Dissection Course Improves the Initially Lower Levels of Visual-Spatial Abilities of Medical Undergraduates. Anat Sci Educ 2020;13(3):333-342.

Hop MJ, Moues CM, **Bogomolova K**, Nieuwenhuis MK, Oen IM, Middelkoop E, Breederveld RS, van Baar ME. Photographic assessment of burn size and depth: reliability and validity. J Wound Care 2014;23(3):144-5, 148-52.

# P

## PhD Portfolio

# PhD PORTFOLIO

| 1. PhD training | Year | Workload |
|---|---|---|
| **General academic skills** | | |
| NIHES Master of Clinical Epidemiology | 2017-2018 | 70 ECTS |
| **General courses** | | |
| Biomedical English Writing and Communication | 2017 | 3 ECTS |
| Endnote, Pubmed and other databases, Medical Library | 2017 | 30 hours |
| PhD introductory meeting | 2018 | 8 hours |
| Data management and data stewardship | 2018 | 12 hours |
| Creative thinking techniques for PhDs | 2018 | 14 hours |
| Personal effectivity and communication | 2018 | 24 hours |
| Basiscursus Regelgeving en Organisatie voor Klinisch onderzoekers (eBROK) | 2020 | 1.5 ECTS |
| Critical Choices in Qualitative Research | 2019 | 2 ECTS |
| **(Inter) national presentations** | | |
| Wetenschappelijke vergadering NVPC, Rotterdam, The Netherlands – oral presentation | 2017 | 20 hours |
| Symposium in innovaties in plastisch chirurgisch leren, Amsterdam, The Netherlands – oral presentation | 2017 | 20 hours |
| An International Association for Medical Education (AMEE) – poster with oral presentation | 2017 | 20 hours |
| Nederlandse Vereniging voor Medisch Onderwijs (NVMO), Egmond aan Zee, The Netherlands – oral presentation | 2018 | 20 hours |
| An International Association for Medical Education (AMEE), Basel, Switzerland – oral presentation | 2018 | 20 hours |
| NVMO PhD day, Utrecht, The Netherlands – oral presentation | 2018 | 10 hours |
| International Federation of Associations of Anatomists (IFAA), London, UK – oral presentation | 2019 | 20 hours |
| NVMO PhD day, Utrecht, The Netherlands – oral presentation | 2019 | 10 hours |
| An International Association for Medical Education (AMEE), Vienna, Austria - oral presentation | 2019 | 20 hours |

| | | |
|---|---|---|
| Nederlandse Vereniging voor Medisch Onderwijs (NVMO), Rotterdam, The Netherlands – oral presentation | 2019 | 20 hours |
| International conference on Residency Education (ICRE), Canada – accepted abstract for oral presentation | 2020 | 3 hours |
| Nederlandse Vereniging voor Heelkunde (NHvH), The Netherlands – accepted abstract for oral presentation | 2020 | 3 hours |

| **Invited lectures** | | |
|---|---|---|
| LUMC LEARN, Leiden, The Netherlands | 2018 | 10 hours |
| LUMC LEARN, Leiden, The Netherlands | 2019 | 10 hours |
| LUMC Groot Onderwijsoverleg, Leiden, The Netherlands | 2019 | 10 hours |
| External visit from department of Anatomy, Leeds University, London, UK | 2019 | 10 hours |
| LUMC Board of directors, Leiden, the Netherlands | 2019 | 10 hours |
| Annual meeting of Experimental Biology – symposium on Visual-Spatial abilities | 2020 | 15 hours |

## 2. Teaching activities

| **Supervising** | | |
|---|---|---|
| Supervision minor Head & Neck reconstructive Surgery, Erasmus MC | 2017 | 10 hours |
| Supervision halve minor Medical Education, LUMC | 2018 - 2020 | 30 hours |
| Master Thesis (Judith Cueto Fernandez), Biomedical Engineering, TU Delft | 2019 - 2020 | 60 hours |

| **Lecturing** | | |
|---|---|---|
| Anatomy of the hand and arm (3rd year students), Erasmus MC | 2017 | 5 hours |
| Academische & Wetenschappelijke Vorming- (1st year students), LUMC | 2020 | 15 hours |

# Curriculum Vitae

# CURRICULUM VITAE

Katerina Bogomolova was born in Tashkent, Uzbekistan, on 26 August 1987. Together with her parents and older sister, she moved to the Netherlands when she was 14 years old in 2001. After a relatively short transition through various levels of secondary education (VMBO, HAVO and VWO) Katerina successfully graduated from the Vechtdal College in Hardenberg in 2006.

From 2006 to 2010 Katerina studied Life Sciences & Innovation Management at the Utrecht University. Finally in 2009, she was admitted to Medicine at the Erasmus University Medical Center in Rotterdam.

During her Medicine study in 2015, Katerina initiated and conducted an educational research project under supervision of prof. dr. Steven Hovius, dr. Jan Sluimers and dr. Eddy Putranto at the department of Plastic and Reconstructive Surgery in Medan, Indonesia. During this project she was inspired by the potentials of new technologies to improve medical education. After obtaining her medical degree, Katerina started her PhD project on the role of 3D learning in anatomical and surgical education in 2017 under supervision of prof. dr. Steven Hovius at the department of Plastic and Reconstructive Surgery at Erasmus MC Rotterdam. In the same year, she attained a master's degree in Clinical Epidemiology at the Netherlands Institute for Health Sciences (NIHES). In 2018, she continued her PhD project resulting in this thesis under supervision of prof. dr. Jos van der Hage at the department of Surgery at the Leiden University Medical Center.

After finishing her PhD in January 2021, Katerina worked as a resident not in training (ANIOS) at the department of Surgery in the Ikazia Hospital, Rotterdam. In November 2021, she started as ANIOS at the department of Plastic and Reconstructive Surgery and Hand Surgery at the Amsterdam University Medical Center. It is her dream to pursue a carrier in the field of Plastic and Reconstructive Surgery and Hand Surgery.

D

# Dankwoord

# DANKWOORD

Beste Jos, professor van der Hage, jij hebt mij de kans gegeven om mijn promotietraject te kunnen volbrengen in Leiden waarvoor mijn grote dank. De vrijheid die jij mij hebt gegeven binnen het onderzoek was voor mij van onschatbare waarde, je openheid voor nieuwe ideeën heeft geleid tot vruchtbare samenwerkingen die onze onderzoeken tot een hoger niveau hebben getild.

Beste Steven, professor Hovius, promoveren was voor mij meer dan alleen onderzoek doen, het heeft mijn blik verruimd en extra handvaten gegeven om mijn vleugels verder uit te slaan, en dat heb ik van jou mogen leren. Bedankt voor het vertrouwen in mij, je steun, je kritische blik die mij continue scherp hield en inspirerende gesprekken waar ik keer op keer wijsheid uit put.

Beste Beerend, dr. Hierck, een betere copromotor kon ik niet wensen - jij stond altijd voor mij klaar, samen waren wij een team waarbij jouw persoonlijke bijdrage en creaties onmisbaar zijn geweest voor ons onderzoek. Dank voor je aanstekelijke enthousiasme en de vele leuke en leerzame momenten waarop wij eindeloos konden sparren over het onderzoek en het leven.

Beste dr. Sluimers, zonder u had dit avontuur niet plaats gevonden. Met u heb ik de halve wereld afgereisd! Vanaf het opstarten van het Google Glass project in Nederland en het uitrollen in Indonesië tot aan onze congres- en operabezoeken in Kiev en de zoektocht naar nieuwe smart glasses in Antwerpen. Daarnaast stond uw eigen deur altijd open voor mij. Bedankt voor uw steun, gezelligheid en kennis die u graag met mij deelt. Uw persoonlijke bijdrage aan onze Z-plastiek studie waardeer ik enorm. Beste Melanie, ik hou nog altijd warme herinneringen over aan onze avonturen in Indonesië.

Mijn lieve mede-promovendi, Belinda, Renée, Marjolein en Kirsten, bedankt voor jullie warme welkom in Leiden. Charlotte, na jouw komst waren wij compleet. Het was een fantastische tijd. Onze kritische feedback op elkaars werk vond ik erg waardevol. Dankzij jullie steun vielen de hobbels mee en leken de dalen helemaal niet zo diep achteraf.

Beste CRIME onderzoeksgroep, bedankt voor het veilige leerklimaat waarin wij als jonge onderzoekers ons hebben kunnen ontwikkelen. Marchien, dank voor je betrokkenheid en steun gedurende het promotietraject.

Ineke van der Ham, Walter van den Broek, Micha Holla, Theo Wiggers en Jeroen van Merriënboer, bedankt voor de vruchtbare samenwerking. Dear professor Bruce Wainman, it has been an honor working with you, thank you for being open to our collaboration. Mary Dankbaar, bedankt voor de fijne supervisie in het eerste jaar en je kritische feedback op mijn werk. Marc Vorstenbosch, voor mij ben jij het levend voorbeeld van 'waar een wil is, is een weg'. Tahmina, met jou heb ik geleerd dat samen onderzoek doen vele malen leuker is dan alleen.

Oud-collega's van de 15e in Rotterdam, ik heb genoten van jullie gezelligheid en ben nog altijd dankbaar voor jullie deelname aan mijn Google Glass experimenten. Stephanie en Jaap, zonder jullie had ik nooit van de doodenge blauwe piste af kunnen komen! Bedankt voor jullie eindeloze geduld, betere skileraren kon ik niet wensen.

Mijn lieve vrienden, jullie maken het leven een stuk gezelliger. Bedankt dat jullie er voor me zijn.

Manon, Chris en Woj, samen zaten wij als geneeskundestudenten in de collegebanken, samen stonden wij als coassistenten op OK de belangrijkste festivals te bespreken en samen vierden wij de door ons bemachtigde PhD plekken. Bedankt voor de mooie momenten samen.

Lieve Steffie, Tal en Gijtje, wat gaan we goed samen. Gijtje, al bij de eerste versies van mijn plan van aanpak heb ik me gesteund gevoeld door jou. Steffie, jouw oprechte interesse in mijn onderzoek, ondanks dat jij niet bekend bent met de wetenschap, waardeer ik enorm.

Lieve Jaap, wij weten beiden ons 'eigen pad' standvastig te bewandelen en daar ben ik trots op. Bedankt dat jij vandaag mijn paranimf bent.

Lieve Tal, wie had er vandaag beter naast mij kunnen staan dan jij? Jij hebt mij vanaf dag één tot het einde van mijn promotietraject bijgestaan. Samen gevierd, samen gehuild. Jij geloofde in mijn succes en dat waardeer ik enorm. Ik voel me vereerd dat jij vandaag naast mij wilt staan als paranimf.

Lieve schoonfamilie, bedankt voor jullie oprechte interesse in mijn werk en met name jullie enthousiasme en support. Een betere tweede familie kan ik niet wensen.

Ira, mijn lieve zus, bedankt dat je er altijd voor me bent ondanks de grote afstand, ik ben blij dat we elkaar hebben.

Lieve Steven, met jou aan mijn zij lijkt niets onmogelijk in dit leven. Jij weet mij als geen ander te motiveren en mij de spiegel op het juiste moment voor te houden. Bedankt voor je onvoorwaardelijke steun en geloof in mij, dat jij in mijn leven bent is een verrijking.

Дорогие мама и папа, спасибо за вашу безоговорочную поддержку. Вы подарили мне возможность выбрать свой путь, который я прошла с успехом. Эта работа посвящается вам.