MEDICAL EDUCATION IN REVIEW

Stereoscopic three-dimensional visualisation technology in anatomy learning: A meta-analysis

Katerina Bogomolova[1,2,3] (iD) | Beerend P. Hierck[2,4,5,6] | Agnes E. M. Looijen[3] | Johanne N. M. Pilon[3] | Hein Putter[6] | Bruce Wainman[7] (iD) | Steven E. R. Hovius[8] | Jos A. van der Hage[1,2]

[1]Department of Surgery, Faculty of Medicine, Leiden University Medical Center, Leiden, The Netherlands

[2]Faculty of Medicine, Center for Innovation of Medical Education, Leiden University Medical Center, Leiden, The Netherlands

[3]Faculty of Medicine, Institute for Medical Education Research Rotterdam, Rotterdam Erasmus University Medical Center, Rotterdam, The Netherlands

[4]Department of Anatomy and Embryology, Faculty of Medicine, Leiden University Medical Center, Leiden, The Netherlands

[5]Leiden Teachers' Academy, Leiden University, Leiden, The Netherlands

[6]Department of Medical Statistics and Bioinformatics, Faculty of Medicine, Leiden University Medical Center, Leiden, The Netherlands

[7]Department of Pathology and Molecular Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada

[8]Department of Plastic and Reconstructive Surgery, Faculty of Medicine, Radboud University Medical Center, Nijmegen, The Netherlands

Correspondence
Katerina Bogomolova, Department of Surgery, Leiden University Medical Center, PO Box 9600, Zone D6-P, 2300 RC Leiden, The Netherlands.
Email: k.bogomolova@lumc.nl

Abstract

Objectives: The features that contribute to the apparent effectiveness of three-dimensional visualisation 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 randomised 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, for example 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 utilised with non-interactive 3D images (ES = −0.87, 95% CI −2.09-0.35; $P = .16$).

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

1 | INTRODUCTION

Three-dimensional visualisation 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 (eg augmented and virtual reality) and two-dimensional (2D) environments (eg 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 lower visual-spatial abilities.[1-5] It has been hypothesised that digital multiple view-based, or 3D, images are being memorised 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 visual-spatial abilities are able to devote more cognitive resources to building mental connections. Students with lower visual-spatial abilities, 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 lower visual-spatial abilities are able to reach the performance level of students with higher visual-spatial abilities.[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 eyes and this can be obtained in 3DVT by presenting slightly shifted 2D images to the eyes.[10] Stereovision can be produced with supportive devices such autostereoscopic displays (eg Alioscopy 3D Display [Alioscopy, Paris, FR]), anaglyphic or polarised glasses, or by a head-mounted display (eg 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. Oculus Rift™ and HTC VIVE™ are predominantly used to create virtual reality environments. A binocular vision of the viewer, though, is required to perceive visual-spatial 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, colouring, 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 3D object results. Making a distinction between stereoscopic and monoscopic 3D visualisations 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 (eg 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 by contrast with 3D view in dissection including haptic feedback) or instructional method (self-regulated learning by contrast with 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 favour 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 visual-spatial ability
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 (ie studies with comparisons made within one single level of instructional design).[13]

2 | 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]

2.1 | 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. The entire search strategy can be found in Appendix S1.

2.2 | 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 randomised controlled trials (RCT) and non-randomised comparative studies. *Studies with a single group with pre-test and post-test, single group post-test 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.

2.3 | Data extraction

Reviewers extracted the following data from each eligible study using a piloted extraction sheet: type of study design, target group and field, inclusion or exclusion criteria (assessment of stereoscopic vision and visual-spatial abilities), 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 categorised

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 memorise 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 memorisation and recall, were assigned as low-order questions. High-order questions were defined as transformation and application of acquired knowledge, including understanding of spatial organisation, 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 memorisation and recall. When information about the type of questions was insufficient, reviewer (KB) requested this information from authors via e-mail.

2.4 | 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.

2.5 | Data analysis

A descriptive analysis was used to summarise the included studies and to describe the effect of visual-spatial abilities 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 (eg the ability of studies to

isolate stereopsis as the only true manipulating element). This resulted in three types of comparisons:

2.5.1 | 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 (eg adjustment of the size, rotation of the model) and/or dynamic exploration (eg walking around the model) in case of interactive augmented and virtual reality environments.

2.5.2 | 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.

2.5.3 | 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 low- and high-order knowledge domains separately.

The standardised 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-post-test design, we used post-test means. Heterogeneity between studies was quantified by I^2 statistics.[32] In case of large inconsistency, for example $I^2 > 50\%$, 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.

3 | RESULTS

3.1 | 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-randomised study design.[34-36] Ultimately, 13 studies were included in the meta-analysis.

3.2 | Study characteristics

Among 16 studies included in the qualitative synthesis, study designs included randomised controlled trials with pre- and post-test (n = 7; 43.8%), post-test only (n = 7; 43.8%) and non-randomised comparative studies (n = 2; 12.4%) (Appendix S2). 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%), behavioural 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%), polarising 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, that is the images was perceived in 3D without head-mounted device (n = 1; 5.9%).

3.3 | Study appraisal

The appraisal of methodological quality and the assessment of risk of bias of the included studies are summarised in Table A3 (Appendix S3). The mean MERSQI score for the 16 included studies was 13.0,

FIGURE 1 Flow diagram of study selection
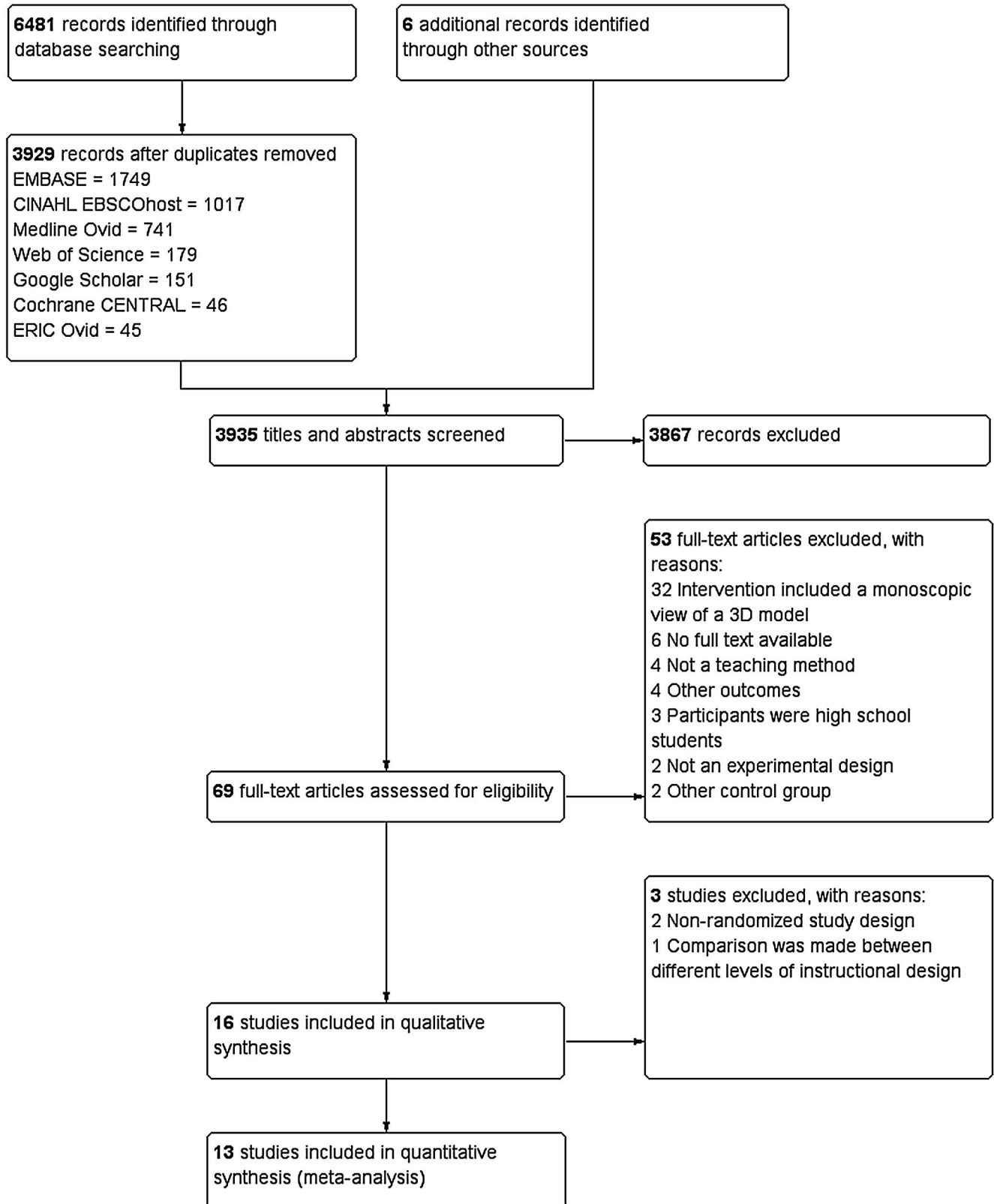
ranging from 10.5 to 15. Reduction in scores and assigned 'high risk' of bias was primarily due to non-randomised study design of several studies. The MERSQI score for the studies included in the meta-analysis ranged from 12.5 and 15.0, with the highest score and no assigned 'high risk' of bias for the comparison 'interactive stereoscopic 3D model versus interactive monoscopic 3D model'.

3.4 | Meta-analysis

3.4.1 | 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 (ie identification task and localisation task),[2] and one study evaluated two interventions (ie virtual reality and interactive augmented reality) separately.[42] A significant positive effect on overall anatomical knowledge was observed in favour of interactive stereoscopic 3D models (ES = 0.53, 95% CI 0.26-0.80; P < .00001; I^2 = 51%; n = 8) (Figure 2). 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 favour 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.

3.4.2 | 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 favour 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 favour 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).

3.4.3 | 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 favour 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]

3.5 | The effect of visual-spatial ability

Visual-spatial abilities 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 visual-spatial abilities by treating it as a confounder and reported that MRT scores significantly predicted performance (r = .548, P < .0001).[42] Five studies evaluated the possible modifying effect of visual-spatial abilities by including visual-spatial abilities in a linear regression analysis as an interaction term[2,37,38] or by stratifying outcomes by visual-spatial abilities.[3,4] A significant interaction caused by visual-spatial abilities was reported in three studies.[3,4,38] This interaction meant that participants with lower visual-spatial abilities benefited significantly more from the stereoscopic view of anatomy than individuals with higher visual-spatial abilities. Moreover, in the 3D stereoscopic group, differences between students with lower and higher visual-spatial abilities were no longer significant.[3,4] One study reported similar interaction that

Study or Subgroup	Weight	Std. Mean Difference IV, Random, 95% CI	Std. Mean Difference IV, Random, 95% CI
De Faria et al. 2016	12.0%	1.08 [0.51, 1.64]	
Hackett and Proctor 2018	17.1%	0.76 [0.39, 1.13]	
Luursema et al. 2008 (I)	11.2%	0.77 [0.17, 1.37]	
Luursema et al. 2008 (L)	11.6%	0.18 [−0.40, 0.76]	
Luursema et al. 2017	11.5%	0.01 [−0.57, 0.60]	
Remmele et al. 2018	15.8%	0.40 [−0.01, 0.82]	
Wainman et al. 2019 (AR)	10.8%	0.04 [−0.58, 0.66]	
Wainman et al. 2019 (VR)	10.1%	0.94 [0.28, 1.59]	
Total (95% CI)	**100.0%**	**0.53 [0.26, 0.80]**	

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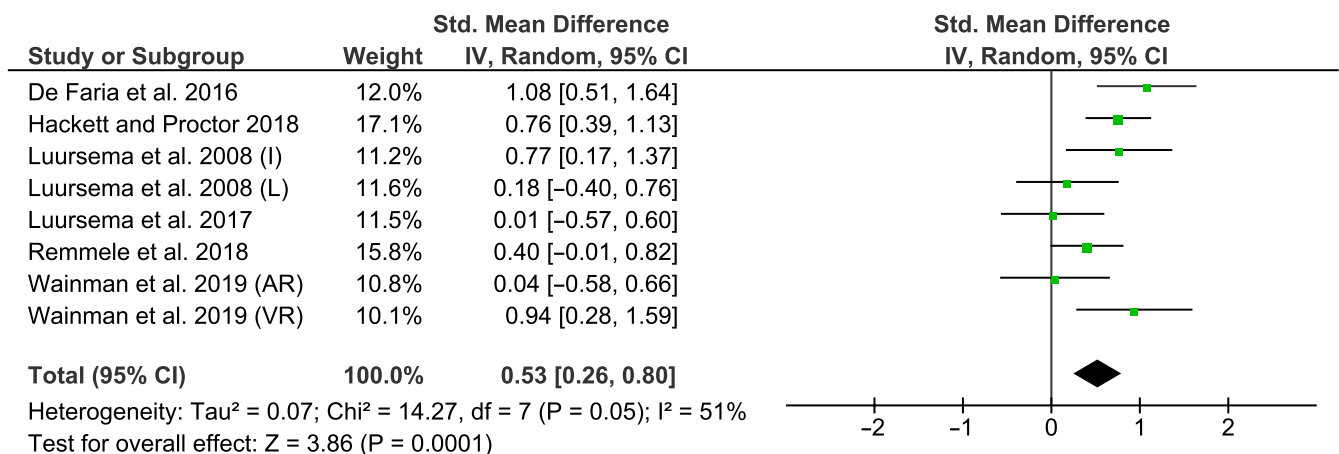Test for overall effect: Z = 3.86 (P = 0.0001)

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

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

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)

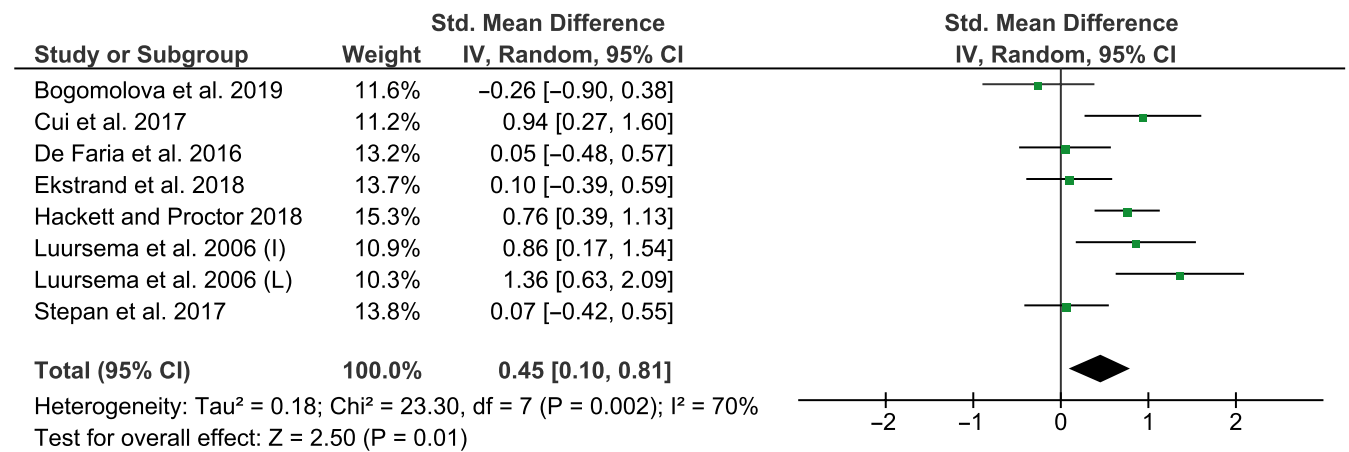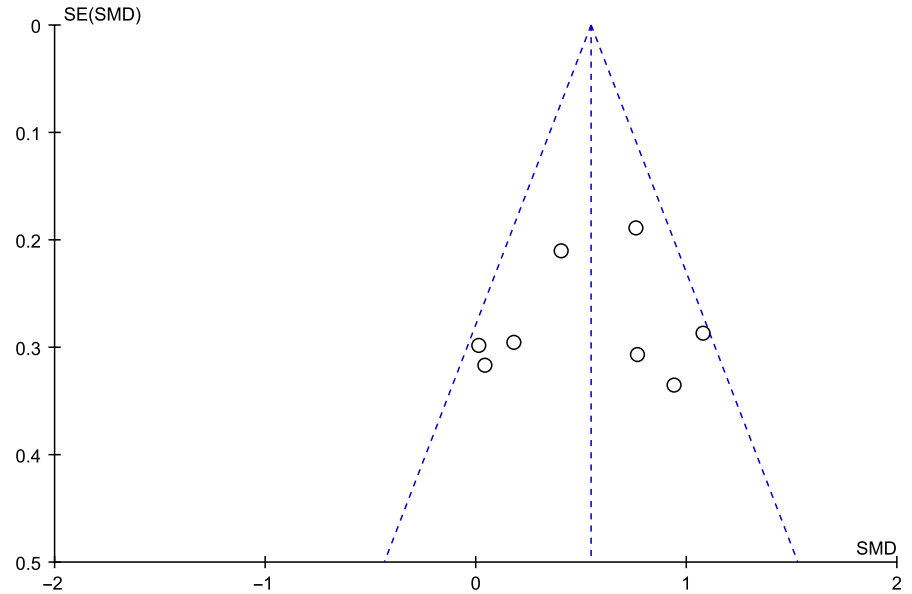| Study or Subgroup	Weight	Std. Mean Difference IV, Random, 95% CI
Bogomolova et al. 2019	11.6%	−0.26 [−0.90, 0.38]
Cui et al. 2017	11.2%	0.94 [0.27, 1.60]
De Faria et al. 2016	13.2%	0.05 [−0.48, 0.57]
Ekstrand et al. 2018	13.7%	0.10 [−0.39, 0.59]
Hackett and Proctor 2018	15.3%	0.76 [0.39, 1.13]
Luursema 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]

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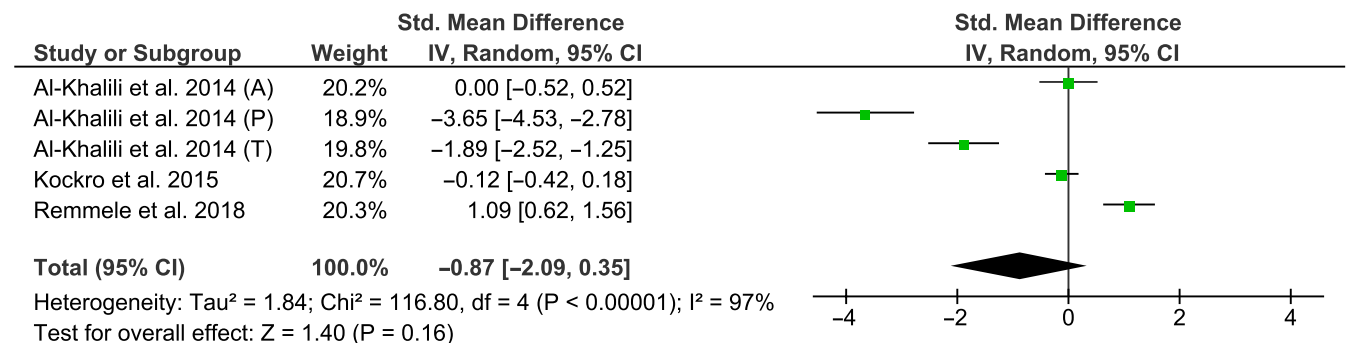

Study or Subgroup	Weight	Std. Mean Difference IV, Random, 95% CI
Al-Khalili et al. 2014 (A)	20.2%	0.00 [−0.52, 0.52]
Al-Khalili et al. 2014 (P)	18.9%	−3.65 [−4.53, −2.78]
Al-Khalili et al. 2014 (T)	19.8%	−1.89 [−2.52, −1.25]
Kockro et al. 2015	20.7%	−0.12 [−0.42, 0.18]
Remmele et al. 2018	20.3%	1.09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

Total (95% CI)   100.0%   −0.87 [−2.09, 0.35]

Heterogeneity: Tau² = 1.84; Chi² = 116.80, df = 4 (P < 0.00001); I² = 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

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 visual-spatial ability could not be performed due to insufficient data.

4 | 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, for example 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 lower visual-spatial abilities, 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 memorised 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 visual-spatial abilities 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 visual-spatial abilities, however, lack this ability and experience difficulties to learn from monoscopic visualisation. This explains why monoscopic 3DVT was found to have disadvantages for students with lower visual-spatial abilities.[3-5,49]

The superiority of stereoscopic 3DVT over monoscopic 3DVT for students with low visual-spatial abilities 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 students with lower visual-spatial abilities. A similar interaction in line of compensating hypothesis has been observed by Berney and colleagues comparing dynamic and static visualisations in learning anatomy.[51] Especially students with lower visual-spatial abilities benefited from learning with dynamic visualisation. This was explained by the required mental representations of the movements that were already provided within dynamic visualisations, while representations within static visualisation still needed to be mentally generated.

Consequently, dynamic visualisation was able to compensate for individuals with lower visual-spatial abilities.

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 with 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 visual spatial ability and visualisation 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 the current review, students with lower visual-spatial abilities benefited most from stereoscopic 3D models and showed a learning trajectory distance from the students with higher visual-spatial abilities. It is important to mention that such interactions are only detectible when outcomes are stratified by the variable (visual-spatial ability) 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 visual-spatial abilities could not be performed due to an insufficient number of studies, these findings suggest that visual-spatial abilities can potentially modify learning outcomes. Such an interaction caused by visual-spatial abilities 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 visual-spatial abilities by the study design (eg randomisation) or statistical analysis (eg including it only as a covariate in the regression analysis). Therefore, for researchers it is essential to make the distinction between various levels of visual-spatial abilities by analysing 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 post-test 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.

4.1 | 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 together. 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 spatial skills. Various levels of complexity could have affected learning outcomes depending on visual-spatial abilities of learners. Third, no distinction was made between different ways of obtaining monoscopic views in control conditions. One study created a monoscopic 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.

5 | 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, visual-spatial abilities 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 utilised in an interactive 3D environment.

CONFLICT OF INTEREST
None.

AUTHORS CONTRIBUTIONS
KB involved in all levels of the project including study conception, methodological design, data collection, data analysis and interpretation and manuscript preparation. BH, AL and AP involved in methodological design, data collection, data interpretation and preparation of the manuscript. HP involved in data analysis and interpretation. BW involved in data analysis, interpretation of the results, preparation and revision of the manuscript. SH and JH involved in study conception, methodological design, interpretation of the results and preparation of the manuscript. All authors approved the final manuscript for submission.

ORCID

Katerina Bogomolova (iD) https://orcid.org/0000-0003-4985-1183
Bruce Wainman (iD) https://orcid.org/0000-0001-7616-961X

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, Ham IJM, Dankbaar MEW, 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.
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, UK: Cambridge University Press; 2001.
9. Hegarty M, Sims VK. Individual differences in use of diagrams as external memory in mechanical reasoning. *Mem Cognit.* 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, eds. *Perception of Space and Motion.* San Diego, CA: 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;32(7):647-652.
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-e34.
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]. Oxford, UK: 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 JAC, Sutton AJ, Ioannidis JPA, 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. *Br J Educ Technol.* 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, Stereopsis KJ. Visuospatial ability, and virtual reality in anatomy learning. *Anat Res Int.* 2017;2017:1-7.
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 JWV, Teixeira MJ, de Moura Sousa Júnior 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, Lamb J, Norman GR. The critical role of stereopsis in virtual and mixed reality learning environments. *Anat Sci Educ*. 2020;13(3):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. *Comput Educ*. 2017;106:150-165.

56. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*, 3rd ed. Philadelphia, PA: 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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.