

The Effectiveness of the Use of Augmented Reality in Anatomy Education: A Systematic Review and Meta-Analysis

Kerem Bölek

Radboud University Nijmegen Medical Centre

Anne-Marie Van Cappellen van Walsum

Radboud University Nijmegen Medical Centre

Guido De Jong

Radboud University Nijmegen Medical Centre

Dylan Henssen (✉ dylan.henssen@radboudumc.nl)

Radboud University Nijmegen Medical Centre

Research Article

Keywords: Augmented Reality, Effectiveness, Meta-analysis, Systematic review

Posted Date: February 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-154748/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on July 27th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-94721-4>.

Abstract

The use of Augmented Reality (AR) in anatomical education has been promoted by numerous authors. Next to financial and ethical advantages, AR has been described to decrease cognitive load while increasing student motivation and engagement. Despite these advantages, the effects of AR on learning outcome varies in different studies and an overview and aggregated outcome on learning anatomy is lacking. Therefore, a meta-analysis on the effect of AR vs. traditional anatomical teaching methods on learning outcome was performed. Systematic database searches were conducted by two independent investigators using predefined inclusion and exclusion criteria. This yielded five papers for meta-analysis totaling 508 participants; 240 participants in the AR-groups and 268 participants in the control groups. (306 females/ 202 males). Meta-analysis showed no significant difference in anatomic test scores between the AR group and the control group (-0.765%; P=0.732). Sub analysis on the use of AR vs. the use of traditional 2D teaching methods showed a significant disadvantage when using AR (-5.685%; P=0.024). Meta-regression analysis showed no significant co-relation between mean difference in test results and spatial abilities (as assessed by the mental rotations test scores). Student motivation and/or engagement could not be included since studies used different assessment tools. This meta-analysis showed that insufficient evidence is present to conclude AR significantly impacts learning outcome and that outcomes are significantly impacted by students' spatial abilities. However, only few papers were suitable for meta-analysis, indicating that there is a need for more well-designed, randomized-controlled trials on AR in anatomy education research.

Introduction

Anatomy education has historically been facilitated by cadavers, anatomical models and drawings in anatomical atlases¹. In line with this, the anatomical assessment is based on the ability to recall spatial relationships between structures, both in two-dimensions (2D) and three-dimensions (3D)². However, with an increasingly cramped curriculum for medical students, anatomy educators have been searching for engaging and interactive teaching methods based on state-of-the-art technologies³. Augmented reality (AR) concerns such a new technologies which is believed to have great potential for anatomy education^{4,5}.

AR has been defined as a technique that allows the user to superimpose virtual objects onto physical objects in real space and allows individuals to interact with both simultaneously. An essential difference with virtual reality concerns that with AR, the user is not completely immersed in a digital environment, which enables the user to combine digital input and real world objects⁶.

Although the research concerning the implementation of AR in anatomical education is relatively limited, there are promising results regarding the teaching potential of AR^{5,7}. Especially with regard to students' motivation to study (neuro)anatomy, various favorable reports have been published over the years⁸⁻¹⁰. The effects of AR on intrinsic anatomy learning have also been investigated by various authors^{11,12}. However, such studies are sparse and more evidence on a meta-study level is needed to investigate whether AR could effectively replace or supplement other anatomy teaching methods. For this reason, the current study set out to perform a systematic literature review and meta-analysis of the available evidence on the impact of AR on learning outcomes in anatomy education.

Materials And Methods

Search strategy and data inclusion

The present study focuses on the effectiveness of learning (neuro) anatomy by students by use of AR and was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines¹³. To assess a wide number of eligible papers, various databases (e.g., Pubmed, Embase, ERIC (Education Resources Information Center), The Cochrane Library, Google Scholar) were searched systematically after an independent librarian was consulted. Searches were conducted until January 2021. Search strings per database are provided in the Supplementary files. There was no restriction in the search strategy on publication data. Additionally, the authors (K.B. and D.H.) hand-searched the reference lists of relevant systematic reviews and included papers. One of the authors (D.H.) contacted corresponding authors of papers when data was missing or when clarification was needed. Selection of relevant articles was carried out by two researchers independently (K.B. and D.H.). The papers eligible for inclusion were original research reports of a comparative study in which the research aim was to investigate the effects of AR on post-intervention anatomic knowledge in university-level human (neuro)anatomical education. These effects needed to be evaluated by any other form of anatomical education (e.g., dissection, atlas-based learning etc.). Case reports, editorial commentaries, systematic or narrative reviews and articles that did not meet the inclusion criteria were excluded.

The first round of assessment of the obtained papers concerned screening title and/or abstract. The second round of assessment comprised full-text assessment and included whether these articles met the aforementioned inclusion criteria to be included. When in disagreement, a third investigator (G.d.J.) was contacted to make the final decision. The PRISMA flow diagram can be appreciated in Fig. 1.

After inclusion, data were extracted from the individual papers using a data extraction sheet by two authors independently (K.B. and D.H.). These data included: 1) type of AR used in the study, 2) type of anatomical education in the control group, 3) number of participants, 4) characteristics of the included participants (i.e., sex, age, study direction), 5) type of anatomical test, 6) mean post-intervention anatomic test scores for the experimental (AR) group, 7) mean post-intervention anatomic test scores for the control group and 8) Mental Rotations Test (MRT) scores in percentages of each

included group as this test assesses the spatial abilities of participants. When the design of the study was a multiple group comparison study, each individual group that was not using AR was considered a separate control group. All control groups were then included for the meta-analysis.

Quality assessment and risk of bias

The quality of the evidence of the studies was graded by two authors independently (K.B. and D.H.) according to the GRADE approach guidelines defined by The Cochrane Collaboration's Handbook¹⁴. Additionally, risk of bias was assessed by two authors independently (K.B. and D.H.). Discrepancies were resolved by discussion or reference to a third author (G.d.J.). Risks of biases which were assessed included: selection bias (criteria 1, 2, 9), performance bias (criteria 3, 4, 10, 11), attrition bias (criteria 6, 7), detection (or measurement) bias (criteria 5, 12) and reporting bias (criterion 8). Also, the Kirkpatrick's model of change of knowledge was assessed for each paper as well. This model evaluates the learning outcomes and classifies these in four levels: 1) reaction; 2A) learning (change in attitude); 2B) learning (modification of knowledge or skills); 3) behavior (change in behavior); 4A) results (change in the system/organizational practice); and 4B) results (improvement in learner performance)^{15,16}. Each potential source of bias was graded as low, high, or unclear. Assessing the risk of bias was performed by the criteria presented in Table 1 following standardized instructions¹⁴.

Table 1
Quality assessment of the evidence provided by the individual papers

Study	Internal validity												Score	Quality	Level in Kirkpatrick's model
	1	2	3	4	5	6	7	8	9	10	11	12			
Moro et al. 2017	+	-	-	-	-	-	+	+	+	+	+	+	60%	Moderate	2B
Barmaki et al. 2019	+	-	-	-	-	+	+	+	+	+	+	-	60%	Moderate	2A, 2B
Bork et al. 2019	-	-	-	-	-	+	+	+	+	+	+	+	60%	Moderate	2A, 2B
Henssen et al. 2019	+	-	-	-	+	+	+	+	+	+	+	+	75%	Moderate	2A, 2B
Bogomolova et al. 2020	+	-	-	-	-	+	+	+	+	+	+	+	75%	Moderate	2A, 2B

1. Was the method of randomization adequate?
2. Was the allocation concealed?
3. Was the participant blinded to the intervention?
4. Was the teacher blinded to the intervention?
5. Was the outcome assessor blinded to the intervention?
6. Was the dropout rate described and acceptable?
7. Were all randomized participants analyzed in the group to which they were allocated?
8. Are reports of the study free of suggestion of selective outcome reporting?
9. Were the groups similar at baseline regarding the most important prognostic indicators?
10. Were co-interventions avoided or similar?
11. Was the compliance acceptable in all groups?
12. Was the timing of the outcome assessment similar in all groups?

+, criterion achieved; -, criterion not achieved; *, assessors initially disagreed

High: >75% of the criteria have been fulfilled [$\geq 10/12$]. Where they have not been fulfilled the conclusions of the study or review are thought very unlikely to have been altered.

Moderate: 50–75% of the criteria have been fulfilled [6–9/12]. Those criteria that have not been fulfilled or not adequately described are thought unlikely to have altered the conclusions.

Low: Less than 50% of the checklist criteria were fulfilled [$< 6/12$]. The conclusions of the study are thought likely or very likely to alter had those criteria been fulfilled⁴⁶⁻⁵⁵.

Levels of change of knowledge according to the model of Kirkpatrick: 1) reaction; 2A) learning (change in attitude); 2B) learning (modification of knowledge or skills); 3) behavior (change in behavior); 4A) results (change in the system/organizational practice); and 4B) results (improvement in learner performance) ^{15,16}

Table 2
Specifications of the included studies and characteristics of the included participants

Study	Anatomy learning task	Type of AR feature	Comparison	Subjects in each group (n)	Mean age (± SD)	Gender (F/M)	Study (MED/ BMS)	Mean test-score in the different groups (%) (± SD)	Mean difference in test-scores	Lower bound – Upper bound (%)
Moro et al. 2017	Studying anatomy of the bones of the skull	1) Tablet-based AR application presenting 3D model of the bones of the skull	2) Headset-based VR application 3) Tablet-based non-AR three dimensional model	1) 17 2) 20 3) 22	1) 19.5 ± 2.3 2) 20.2 ± 3.5 3) 22.2 ± 8.0	1) 7/10 2) 12/8 3) 12/10	N/A	1) 62.5 ± 17.1* 2) 64.5 ± 18.5* 3) 66.5 ± 18.5*	1-2) -2.0% 1-3) -4.0%	1-2) -13.5–9.5% 1-3) -15.1–7.2%
Barmaki et al. 2019	Body painting of musculoskeletal anatomy of the upper and lower limb	1) REFLECT; virtual mirror with augmented anatomical over-projection	2) No REFLECT; virtual mirror without augmented anatomical over-projection	1) 164 2) 124	<i>Total:</i> 19.8 ± 2.0	<i>Total:</i> 178/110	N/A	1) 43.0 ± 28.4 2) 39.2 ± 28.8	1-2) 3.8%	1-2) -2.9–10.5%
Bork et al. 2019	Studying gross anatomy of body parts (pelvis, shoulder, chest, abdomen, and extremities)	1) MagicMirror; virtual mirror with augmented anatomical over-projection	2) Anatomage; a virtual dissection table 3) Traditional, 2D anatomical atlases	1) 24 2) 24 3) 24	<i>Total:</i> 21.4 ± 3.4	<i>Total:</i> 49/23	N/A	1) 56.0 ± 14.1 2) 55.2 ± 11.0 3) 59.1 ± 16.9	1-2) 0.8% 1-3) -3.1%	1-2) -6.3–8.0% 1-3) -11.9–5.7%
Henssen et al. 2019	Studying neuroanatomy	1) GreyMapp ; tablet-based AR application presenting a 3D model of the human brain	2) Cross-sections of the human brain	1) 15 2) 16	1) 19.3 ± 2.3 2) 19.1 ± 0.8	1) 6/9 2) 6/10	1) 13/2 2) 10/6	1) 50.0 ± 10.2 2) 60.6 ± 12.4	1-2) -10.6%	1-2) -18.6–-2.6%
Bogomolova et al. 2020	Studying lower limb anatomy	1) Headset-based AR application	2) Non-AR 3D desktop model 3) Traditional, 2D anatomical atlases	1) 20 2) 20 3) 18	1) 18.5 ± 0.8 2) 18.7 ± 1.0 3) 18.7 ± 0.7	1) 12/8 2) 13/6 3) 11/7	1) 17/3 2) 16/4 3) 14/4	1) 47.8 ± 9.8 2) 38.5 ± 14.3 3) 50.9 ± 13.8	1-2) 9.3% 1-3) -3.1%	1-2) 1.7–16.9% 1-3) -10.8–4.6%

*=Standard deviations were derived from Boxplot analysis

AR = Augmented reality; BMS = Biomedical sciences; F = Female; M= Male; MED = Medicine; N/A= Not available; VR= Virtual reality

Statistical analysis

The statistical package SPSS Statistics, version 25 (IBM Corp., Armonk, NY) was used for descriptive statistical analyses of the aggregated data. Descriptive statistical analyses were represented as mean with ± standard deviation (± SD). Meta-analysis was carried out by use of the visual front-end for the R-package (www.r-project.org; Metafor)¹⁷: OpenMeta[Analyst] software (MetaAnalyst, Tufts Medical Center (Wallace et al., 2012)). A forest-

plot was created to graphically display the estimated differences in pre-intervention and post-intervention test results from the included studies, along with the overall results. In addition, OpenMeta[Analyst] was used to assess heterogeneity. Heterogeneity in meta-analyses refers to the variation in outcomes between included studies. To measure heterogeneity, Cochran's Q was calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies. To improve interpretation, the heterogeneity index (I_2), defined as the proportion of total variability explained by heterogeneity and refers to the percentage of variation across studies, was introduced¹⁸. I_2 is independent from the number of studies included in the meta-analysis. Therefore, I_2 highlights the inconsistency across studies and ranges from 0% (i.e., no heterogeneity) to 100% (i.e., the highest heterogeneity).

Ethical approval

Ethical approval was not applicable for conducting this systematic review and meta-analysis.

Results

Systematic searching and systematic assessment of the retrieved papers resulted in the inclusion of five papers in which AR was compared with another form of anatomical learning, as shown in **Figure 1**.^{11,12,19-21} The assessment for the risk of bias and the level of change of knowledge according to the model of Kirkpatrick is summarized in **Table 1**. See **Table 2** for more information on the participants in the included studies. All papers showed to be of moderate quality with minimal risks of bias.

Study Characteristics

The initial search yielded 430 results found in different databases of which 23 were duplicates and removed. Evaluating the title and abstract, 43 records were chosen to be screened. Of these, 12 papers were eligible for the qualitative synthesis. After evaluating full text, 12 papers were found to match our inclusion criteria, of which 7 proved to be irrelevant to our aim. The 5 remaining papers met the inclusion criteria. However some of the required outcomes, such as student motivation was not reported in all of the papers. The PRISMA flowchart shows the details and the search strategy can be found in the supplementary files. The assessment of the risk of bias was done according to the model of Kirkpatrick and is summarized in **Table 1**. The studies were synthesized by identifying the similar key themes and statements in these papers and then by independent reviews and later consensus building reclassifying these similarities and gathering conclusions from them following the PICO framework.

Participant variation

The total amount of participant were N= 569, of which 306 female. Participants originated from several countries, namely Australia, United States, Germany and the Netherlands. Undergraduates studying anatomy were sought out. The five studies have similar age groups, with the clear outlier of one paper's third group²¹. The means range from 18.5 to 22.5 years of age. Three studies reported the ratio of included biomedical students to medical students^{12,19,20}, which can be seen in **Table 2**. The groups show similarities in age, future academic aims and MRT scores. The effect of MRT scores has been examined in three papers^{11,19,20}. MRT scores showed to have a significant impact on the pre and posttest scores. Bork et al. showed that participants with low MRT scores using AR had higher scores compared to control, which was in accordance with the findings of Bogomolova et al., 2020.

Intervention heterogeneity

The AR interventions show differences in their approach to AR. Henssen et al., 2019 and Moro et al., 2017 shows a practical tablet based 3D model, while two studies opted for virtual mirrors with AR capabilities, called REFLECT^{11,12}. This mirror possess the ability to virtually project musculature on a subject. A headset-based AR application has been used in one study¹⁹. All these interventions conform to the definition of AR. However, the differences should be noted in the form of AR and the implications, such as the adverse events reported by Moro et al, 2017. These showed that AR users experienced more general discomfort in their use compared to tablet users²¹. Henssen et al., 2019 reported that students needed to get used to the device, causing some discomfort. Magic Mirror was claimed to be tiring to use after long learning sessions, according to three participants from Bork et al. 2019 while no such feedback was given in Barmaki et al., 2019. Moreover, no adverse effects were reported by Bogomolova et al., 2020.

Controls

Traditional teaching methods have been used, such as cross-sections and anatomical atlases, by three studies^{11,19,20}. Two of these studies used a virtual dissection table and a non-AR 3D desktop model respectively, while the latter had cross-sections as control. In the study of Barmaki et al. 2019 the virtual mirror without superimposing AR features functioned as control. Moro et al., 2017 compared AR to a VR headset and a conventional tablet based 3D model.

The effects on learning

The primary outcome measure was the effectiveness on learning, measured with the difference in pre- and posttest scores. The tests consisted of multiple choice questions in all of the studies, where some studies opted to supplement the tests with open ended questions, regarding the chosen

anatomical structures. Little to no significant difference were found in the effectiveness on learning anatomy when looking at test scores. Notwithstanding, Bork et al. reported that the AR group did score significantly higher than the virtual dissection table (Anatomage) group. However, no difference between the conventional atlas group and the AR group was found¹¹. Conversely, Barmaki and colleagues found REFLECT users did score significantly higher than their virtual mirror controls¹². MRT scores showed to be of importance as several studies found that students with lower MRT scores learned more with the 3D AR models than with conventional materials.

Secondary outcomes

In the study of Moro et al., 2017 adverse effects were reported for the VR studytool, which caused students to experience nausea, headaches and dizziness. No such symptoms and problems plagued the use of their AR tool. Discomfort was also experienced by students using GreyMapp, as they reported trouble with getting used to operating the application. In combination with taking notes during the lesson, some students assumed uncomfortable positions to multitask. This problem was easily solved by creating a bigger tablet interface. In the REFLECT study, it was reported that time on task increased significantly. In addition, students engagement was significantly higher in the AR group, causing the longer time on task.

Henssen et al. reportedly did not find an increase in motivation when comparing the AR group to the conventional group. However, focus group interviews showed that students did find the concept novel and interesting. Additionally, some students expressed their disappointment with not being able to work with the program²⁰. Engagement was gauged differently in the study of Barmaki et al., 2019, where they measured time on task has been suggested as an important marker for knowledge retention and student engagement. The time on task was significantly higher in the AR group, compared to controls (P=0.01). Finally, a significant difference was found by Bogomolova et al. in the enjoyment during learning between 2D anatomical models and the AR intervention (P=0.003)¹⁹. **Table 2** summarizes the outcomes

Meta-analysis

Meta-analysis showed a substantial heterogeneity in the included papers (Tau²=21.301; Q=15.493; df=7; I₂=54.82%; P=0.030), which complicated further analysis. Based on the mean differences in anatomic test scores (%) between the AR groups and the control groups, a difference of -0.765% was estimated (P=0.732). This indicated that there was no significant advantage or disadvantage when learning anatomy with AR (**Table2; Figure2**). Sub analysis was carried out on studies using 2D anatomy teaching methods as a comparison to AR-based learning^{11,19,20}. This sub analysis showed significant lower mean anatomic test scores for the AR-groups (P=0.024) in studies which showed a low interstudy heterogeneity (Tau²=1.927; Q=2.224; df=2; I₂=10.05%; P=0.329), as seen in **Figure 3**. In order to observe whether outcomes of the different groups (AR vs. control groups) are impacted by spatial abilities of the participants, a meta-regression analysis was performed for the studies that 1) compared AR-features with 2D anatomy teaching methods and 2) used a MRT to assess spatial ability^{11,19,20}. Meta-regression showed no significant relation between mean difference in anatomic test results (%) and mean difference in MRT scores (%) between the AR- and control-groups (Omnibus P=0.229), which can be appreciated in **Figure4**.

Discussion

Although the use of cadavers and/or dissections form the cornerstone of anatomical education for medical and biomedical sciences students, various limitations constrain their use. Therefore, various other teaching options have been explored, including AR. AR is explicitly useful in anatomical education as it presents the first consumer-grade technology that can depict realistic 3D models and concepts to students, which, at the same time, can be directed by a teacher²². However, the present meta-analysis showed that AR yields no significant learning benefits when compared to other forms of anatomical education. Moreover, a significant lower anatomic test score was observed when comparing the results from the AR-groups to groups that used 2D anatomical learning methods (e.g., traditional anatomical atlases and cross-sections). The results from the present meta-analysis partially conflict with the results from the meta-analysis of Yammine and Violato (2015) in which it was found that three dimensional visualization techniques (1) resulted in higher factual knowledge, (2) yielded significant better result in spatial knowledge acquisition, and (3) produced significant increase in user satisfaction and in learners' perception of the effectiveness of the learning tool²³. However, these three dimensional visualization techniques included various 3D images, annotated radiological data and VR simulators and that did not include AR features. On the use of AR in anatomical education, two other recent meta-analyses have been published. The publication of Mori et al. (2020), although also integrating VR methods and non-anatomical education purposes (e.g., physiology education), demonstrated that VR and AR can be used as delivery methods in medical education, without any adverse effects on student performance. Although not supported by their analyses, Mori et al. also expressed that there is a chance that the use of these technologies may have a positive impact on students spatial understanding and 3D comprehension of anatomical structures²⁴. A second meta-analysis, however focusing on VR, showed that VR may act as an efficient way to improve the learners' level of anatomy knowledge²⁵. The present meta-analysis partially contradicts the conclusions of the other studies, showing that AR can indeed worsen the learners' performance when compared to 2D anatomy teaching methods. An explanation for these different outcomes can be explained by the fact that the present study maintained strict inclusion criteria and thereby only focused on the effects of AR in anatomy education. On the one hand, this could have purified the results, whereas on the other hand, this could cause an overestimation of the effects related to a limited sample size.

Spatial ability, cognitive load and the use of AR

One of the co-variables in most studies investigating AR concerns spatial ability. Most studies use the MRT to assess spatial ability of participants. The MRT assesses mental visualization and mental rotation, which are considered the main components of visual-spatial abilities. The MRT concerns a 24-item psychometric questionnaire designed in 1971²⁶ and previously validated by Vandenberg and Kuse (1978)²⁷. The findings of three of the included studies that used MRT^{11,19,20} showed that an aptitude–treatment interaction caused by visual-spatial abilities needs to be considered when reviewing evidence of AR in anatomical learning. However, no significant correlation was found between the mean difference in anatomic test scores and the MRT scores of the different groups in this meta-analysis. This could be due to the fact that only limited data was available. On the contrary, previous studies which focused on spatial ability and the use of 3D visualization methods found that significant differences in pre-intervention spatial ability confounded the study results²⁸⁻³¹. Still, various reports have shown that cognitive load decreases when students study anatomy by use of AR^{20,32}. This could, however, not be incorporated into this meta-analysis as most of the included papers did not provide this information.

Motivation and student engagement

Numerous studies reported improvements in the learners' motivation after implementation of AR in different fields of education³³⁻³⁶. Literature has suggested that AR would be attractive to students, increasing their motivation to learn anatomy³⁷⁻³⁹. Several studies investigated various forms of student motivation with regard to learning anatomy. For example, Allen et al. (2016) reported that students felt confident that learning with 3D models, including AR 3D models, could help them to understand anatomical concepts. Also, the majority of the respondents would encourage the development of similar learning sources⁴⁰. Kucuk et al. distilled from interviewing students that more permanent learning was achieved in a shorter time by using AR⁴¹. Such permanent learning, however, remains rather understudied in research on AR in anatomy education. Another report by our group showed that students feel motivated to study neuroanatomy by use of AR, although men and women and students from different study directions have different attitudes towards learning with AR. As well, students expressed that they felt AR was especially beneficial to study structures that cannot be visualized properly by use of prosected cadavers (i.e., the subcortical structures of the brain)⁴². Although most of the included studies in the present meta-analysis included motivation as a (secondary) outcome measure^{11,12,20,21}, there is still no validated method to measure motivation. Therefore, this could not be included into this meta-analysis. Future research elucidating methods of gathering data on student motivation will therefore provide valuable insights. In addition, the novelty effect, which is defined as "a person's subjective first response to (using) a technological innovation", plays an important role in the studies that used AR as an anatomical teaching method⁴³. Previous studies noted that as the novelty effect wears off, users discontinue their use of new technologies, indicating a loss of interest and motivation^{43,44}. This could partially be explained by the law of diminishing returns, as novel technologies create inherent interest, which tapers off after students get familiarized with their new environments⁴⁵.

Strengths and limitations

One of the strengths of the present meta-analysis concerns the systematic search for available literature and the independent consideration of each paper prior to inclusion and the independent assessment of the risk of bias, level of change in education as defined by Kirkpatrick and the results. A limitation of the present meta-analysis concerns the substantial heterogeneity of the included papers. In addition, testing of anatomical knowledge was performed by using a combination of multiple-choice questions, matching questions and open-ended questions. One of the strengths of the meta-analysis is caused by the consequent use of a validated MRT^{26,27} to assess spatial ability in the included studies. A limitation, on the other hand, is caused by the lack of validated tools to evaluate students' engagement, motivation and cognitive load.

Conclusions

This meta-analysis showed that AR has no significant effects on students' learning anatomy when comparing multiple forms of educational tools. However, when comparing 2D anatomical teaching methods, AR was found to significantly decrease mean anatomic test scores. No significant correlation was found between spatial ability and learning outcome in this meta-analysis. The beneficial characteristics of AR (i.e., lower cognitive load and higher student engagement/motivation) could not be meta-analyzed due to heterogeneity in the measuring methods. Further research that formally measures these parameters is needed to identify these beneficial factors of AR learning in a larger population. In addition, integration of AR in anatomy education needs to be studied thoroughly in order to find the most effective implementation of this technology.

References

1. Estai, M. & Bunt, S. Best teaching practices in anatomy education: A critical review. *Annals of Anatomy-Anatomischer Anzeiger*. **208**, 151–157 (2016).
2. Gonzales, R. A., Ferns, G., Vorstenbosch, M. A. T. M. & Smith, C. F. Does spatial awareness training affect anatomy learning in medical students? *Anat Sci Educ*. **13**, 707–720 (2020).
3. Moro, C., Stromberga, Z. & Birt, J. in *Clinical Education for the Health Professions: Theory and Practice* (eds Debra Nestel, Gabriel Reedy, Lisa McKenna, & Suzanne Gough) 1–22(Springer Singapore, 2020).
4. Kamphuis, C., Barsom, E., Schijven, M. & Christoph, N. Augmented reality in medical education? *Perspect Med Educ*. **3**, 300–311 <https://doi.org/10.1007/s40037-013-0107-7> (2014).
5. Chytas, D. *et al.* The role of augmented reality in Anatomical education: An overview. *Annals of Anatomy-Anatomischer Anzeiger*, 151463(2020).

6. Azuma, R. T. A survey of augmented reality. *Presence-Teleop Virt.* **6**, 355–385 <https://doi.org/10.1162/pres.1997.6.4.355> (1997).
7. Moro, C. *et al.* Virtual and augmented reality enhancements to medical and science student physiology and anatomy test performance: A systematic review and meta-analysis. *Anat Sci Educ* **n/a**, doi:<https://doi.org/10.1002/ase.2049>.
8. Kugelmann, D. *et al.* An Augmented Reality magic mirror as additive teaching device for gross anatomy. *Ann Anat.* **215**, 71–77 <https://doi.org/10.1016/j.aanat.2017.09.011> (2018).
9. Ferrer-Torregrosa, J., Torralba, J., Jimenez, M. A., Garcia, S. & Barcia, J. M. A. R. B. O. O. K. Development and Assessment of a Tool Based on Augmented Reality for Anatomy. *J Sci Educ Technol.* **24**, 119–124 <https://doi.org/10.1007/s10956-014-9526-4> (2015).
10. Ferrer-Torregrosa, J. *et al.* Distance learning icts and flipped classroom in the anatomy learning: comparative study of the use of augmented reality, video and notes. *Bmc Med Educ* **16**, doi:ARTN 230 1186/s12909-016-0757-3 (2016).
11. Bork, F. *et al.* The Benefits of an Augmented Reality Magic Mirror System for Integrated Radiology Teaching in Gross Anatomy. *Anat Sci Educ.* **12**, 585–598 <https://doi.org/10.1002/ase.1864> (2019).
12. Barmaki, R. *et al.* Enhancement of Anatomical Education Using Augmented Reality: An Empirical Study of Body Painting. *Anat Sci Educ.* **12**, 599–609 <https://doi.org/10.1002/ase.1858> (2019).
13. Moher, D. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Rev Esp Nutr Hum Die.* **20**, 148–160 <https://doi.org/10.14306/renhyd.20.2.223> (2016).
14. Higgins, J. P. *et al.* Cochrane handbook for systematic reviews of interventions. (John Wiley & Sons 2019).
15. Kirkpatrick, D. Evaluating training programs: The four levels. (Berrett-Koehler Publishers 1994).
16. Steinert, Y. *et al.* A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach.* **28**, 497–526 <https://doi.org/10.1080/01421590600902976> (2006).
17. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw.* **36**, 1–48 <https://doi.org/10.18637/jss.v036.i03> (2010).
18. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *Brit Med J.* **327**, 557–560 <https://doi.org/10.1136/bmj.327.7414.557> (2003).
19. Bogomolova, K. *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.* <https://doi.org/10.1002/ase.1941> (2020).
20. Henssen, D. *et al.* Neuroanatomy Learning: Augmented Reality vs. Cross-Sections. *Anat Sci Educ.* <https://doi.org/10.1002/ase.1912> (2019).
21. Moro, C., Stromberga, Z., Raikos, A. & Stirling, A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anat Sci Educ.* **10**, 549–559 <https://doi.org/10.1002/ase.1696> (2017).
22. Turney, B. W. Anatomy in a modern medical curriculum. *Ann R Coll Surg Engl.* **89**, 104–107 <https://doi.org/10.1308/003588407X168244> (2007).
23. Yammine, K. & Violato, C. A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. *Anat Sci Educ.* **8**, 525–538 <https://doi.org/10.1002/ase.1510> (2015).
24. Moro, C. *et al.* Virtual and augmented reality enhancements to medical and science student physiology and anatomy test performance: A systematic review and meta-analysis. *Anat Sci Educ.* <https://doi.org/10.1002/ase.2049> (2020).
25. Zhao, J., Xu, X., Jiang, H. & Ding, Y. The effectiveness of virtual reality-based technology on anatomy teaching: a meta-analysis of randomized controlled studies. *BMC Medical Education.* **20**, 127 <https://doi.org/10.1186/s12909-020-1994-z> (2020).
26. Shepard, R. N. & Metzler, J. Mental Rotation of 3-Dimensional Objects. *Science.* **171**, 701–701 <https://doi.org/10.1126/science.171.3972.701> (1971).
27. Vandenberg, S. G. & Kuse, A. R. Mental Rotations, a Group Test of 3-Dimensional Spatial Visualization. *Percept Motor Skill.* **47**, 599–604 <https://doi.org/10.2466/pms.1978.47.2.599> (1978).
28. Garg, A., Norman, G. R., Spero, L. & Maheshwari, P. Do virtual computer models hinder anatomy learning? *Acad Med.* **74**, S87–S89 (1999). Doi 10.1097/00001888-199910000-00049
29. Garg, A. X., Norman, G. & Sperotable, L. How medical students learn spatial anatomy. *Lancet.* **357**, 363–364 [https://doi.org/10.1016/S0140-6736\(00\)03649-7](https://doi.org/10.1016/S0140-6736(00)03649-7) (2001).
30. Garg, A. X., Norman, G. R., Eva, K. W., 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.* **77**, S97–S99 (2002). Doi 10.1097/00001888-200210001-00030
31. Levinson, A. J., Weaver, B., Garside, S., McGinn, H. & Norman, G. R. Virtual reality and brain anatomy: a randomised trial of e-learning instructional designs. *Med Educ.* **41**, 495–501 <https://doi.org/10.1111/j.1365-2929.2006.02694.x> (2007).
32. Kucuk, S., Kapakin, S. & Goktas, Y. Learning anatomy via mobile augmented reality: Effects on achievement and cognitive load. *Anat Sci Educ.* **9**, 411–421 <https://doi.org/10.1002/ase.1603> (2016).
33. Di Serio, A., Ibanez, M. B. & Kloos, C. D. Impact of an augmented reality system on students' motivation for a visual art course. *Comput Educ.* **68**, 586–596 <https://doi.org/10.1016/j.compedu.2012.03.002> (2013).
34. Jara, C. A., Candelas, F. A., Puente, S. T. & Torres, F. Hands-on experiences of undergraduate students in Automatics and Robotics using a virtual and remote laboratory. *Comput Educ.* **57**, 2451–2461 <https://doi.org/10.1016/j.compedu.2011.07.003> (2011).

35. Liu, T. Y. & Chu, Y. L. Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Comput Educ.* **55**, 630–643 <https://doi.org/10.1016/j.compedu.2010.02.023> (2010).
36. Iwata, T., Yamabe, T. & Nakajima, T. Augmented Reality Go: Extending Traditional Game Play with Interactive Self-Learning Support. *Ieee Int Conf Embed.* 105–114 <https://doi.org/10.1109/Rtcsa.2011.43> (2011).
37. Lee, K. Augmented Reality in Education and Training. *Techtrends.* **56**, 13–21 (2012). DOI 10.1007/s11528-012-0559-3
38. Shen, R. M., Wang, M. J. & Pan, X. Y. Increasing interactivity in blended classrooms through a cutting-edge mobile learning system. *Brit J Educ Technol.* **39**, 1073–1086 <https://doi.org/10.1111/j.1467-8535.2007.00778.x> (2008).
39. Huang, Y. M., Lin, Y. T. & Cheng, S. C. Effectiveness of a Mobile Plant Learning System in a science curriculum in Taiwanese elementary education. *Comput Educ.* **54**, 47–58 <https://doi.org/10.1016/j.compedu.2009.07.006> (2010).
40. Allen, L. K., Eagleson, R. & de Ribaupierre, S. Evaluation of an online three-dimensional interactive resource for undergraduate neuroanatomy education. *Anat Sci Educ.* **9**, 431–439 <https://doi.org/10.1002/ase.1604> (2016).
41. Kucuk, S., Kapakin, S. & Goktas, Y. Learning Anatomy via Mobile Augmented Reality: Effects on Achievement and Cognitive Load. *Anat Sci Educ.* **9**, 411–421 <https://doi.org/10.1002/ase.1603> (2016).
42. Bölek, K. A., De Jong, G., Van der Zee, I., Van Cappellen, A. M. & Henssen, D. J. H. A. Mixed-methods exploration of students' motivation in using augmented reality in neuroanatomy education with prosected specimens. *SUBMITTED*(2020).
43. Sung, J., Christensen, H. I. & Grinter, R. E. in Proceedings of the 4th ACM/IEEE international conference on Human robot interaction 45–52(Association for Computing Machinery, La Jolla, California, USA, 2009).
44. Mutsuddi, A. U. & Connelly, K. in 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops. 33–40.
45. Stebbins, J. The law of diminishing returns. *Science.* **99**, 267–271 <https://doi.org/10.1126/science.99.2571.267> (1944).
46. Guyatt, G. H. *et al.* GRADE guidelines 6. Rating the quality of evidence—imprecision. *Journal of clinical epidemiology.* **64**, 1283–1293 <https://doi.org/10.1016/j.jclinepi.2011.01.012> (2011).
47. Guyatt, G. H. *et al.* GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *Journal of clinical epidemiology.* **64**, 1294–1302 <https://doi.org/10.1016/j.jclinepi.2011.03.017> (2011).
48. Guyatt, G. H. *et al.* GRADE guidelines: 5. Rating the quality of evidence—publication bias. *Journal of clinical epidemiology.* **64**, 1277–1282 <https://doi.org/10.1016/j.jclinepi.2011.01.011> (2011).
49. Guyatt, G. H. *et al.* GRADE guidelines: 8. Rating the quality of evidence—indirectness. *Journal of clinical epidemiology.* **64**, 1303–1310 <https://doi.org/10.1016/j.jclinepi.2011.04.014> (2011).
50. Guyatt, G. H. *et al.* GRADE guidelines: 9. Rating up the quality of evidence. *Journal of clinical epidemiology.* **64**, 1311–1316 <https://doi.org/10.1016/j.jclinepi.2011.06.004> (2011).
51. Guyatt, G. H. *et al.* GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *Journal of clinical epidemiology.* **64**, 407–415 <https://doi.org/10.1016/j.jclinepi.2010.07.017> (2011).
52. Balshem, H. *et al.* GRADE guidelines: 3. Rating the quality of evidence. *Journal of clinical epidemiology.* **64**, 401–406 <https://doi.org/10.1016/j.jclinepi.2010.07.015> (2011).
53. Guyatt, G. *et al.* GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *Journal of clinical epidemiology.* **64**, 383–394 <https://doi.org/10.1016/j.jclinepi.2010.04.026> (2011).
54. Guyatt, G. H. *et al.* GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of clinical epidemiology.* **64**, 395–400 <https://doi.org/10.1016/j.jclinepi.2010.09.012> (2011).
55. Higgins, J. P. *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj.* **343**, d5928 <https://doi.org/10.1136/bmj.d5928> (2011).

Figures



PRISMA 2009 Flow Diagram

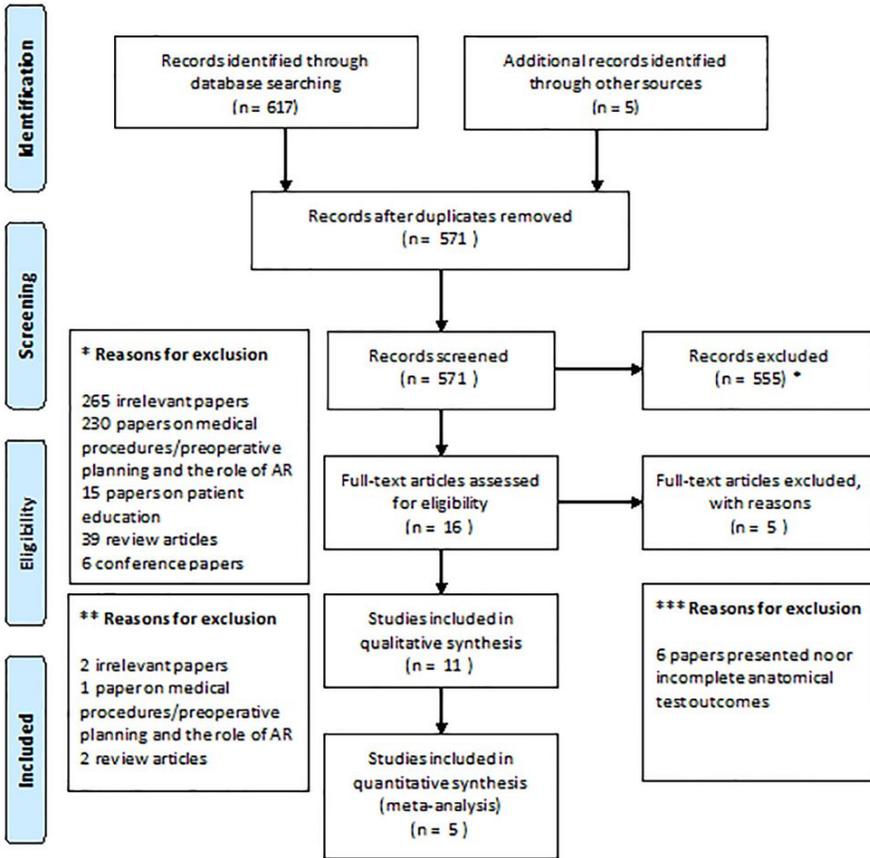


Figure 1
PRISMA flow diagram for the systematic review detailing the database searches, the number of abstracts screened and the full texts retrieved

Studies	Estimated difference (%)	(95%-CI)
Moro et al. (1) 2017	-2.000	(-13.481, 9.481)
Moro et al. (2) 2017	-4.000	(-15.218, 7.218)
Barmaki et al. 2019	3.200	(-5.688, 12.088)
Bork et al. (1) 2019	0.840	(-6.301, 7.981)
Bork et al. (2) 2019	-3.110	(-11.907, 5.687)
Henssen et al. 2019	-10.600	(-18.573, -2.627)
Bogomolova et al. (1) 2020	9.300	(1.702, 16.898)
Bogomolova et al. (2) 2020	-3.100	(-10.787, 4.587)
Overall (I²=51.85 %, P=0.042)	-0.987	(-5.386, 3.412)

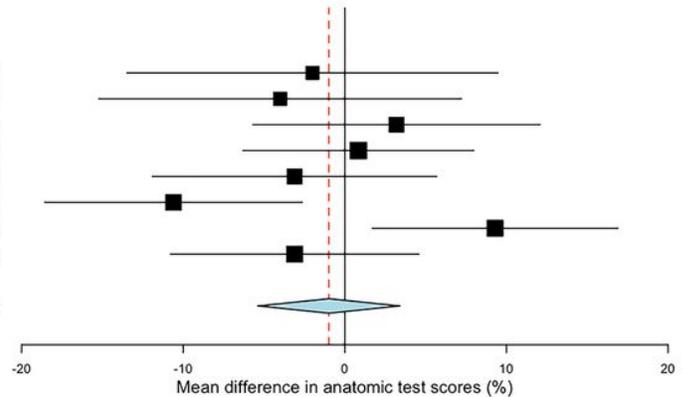


Figure 2
Forest plot showing the estimated mean difference in anatomic test scores (%) from the different included studies investigating AR as compared with other forms of anatomical education AR= Augmented reality; 95%-CI= 95%-Confidence interval

Studies	Estimated difference (%)	(95%-CI)
Bork et al. (2) 2019	-3.110	(-11.907, 5.687)
Henssen et al. 2019	-10.600	(-18.573, -2.627)
Bogomolova et al. (2) 2020	-3.100	(-10.787, 4.587)
Overall (I²=10.05 %, P=0.329)	-5.685	(-10.629, -0.742)

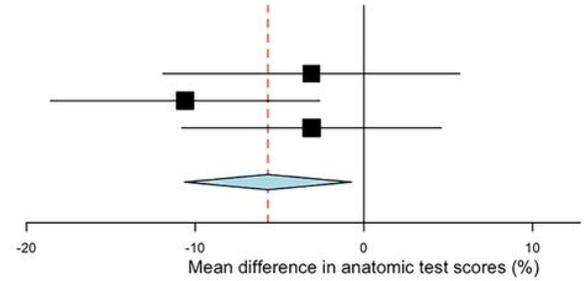


Figure 3
 Forest plot showing the estimated mean difference in anatomic test scores (%) from the included studies addressing AR vs. 2D forms of anatomical education (i.e., traditional anatomical atlases, radiological data) AR= Augmented reality; 95%-CI= 95%-Confidence Interval

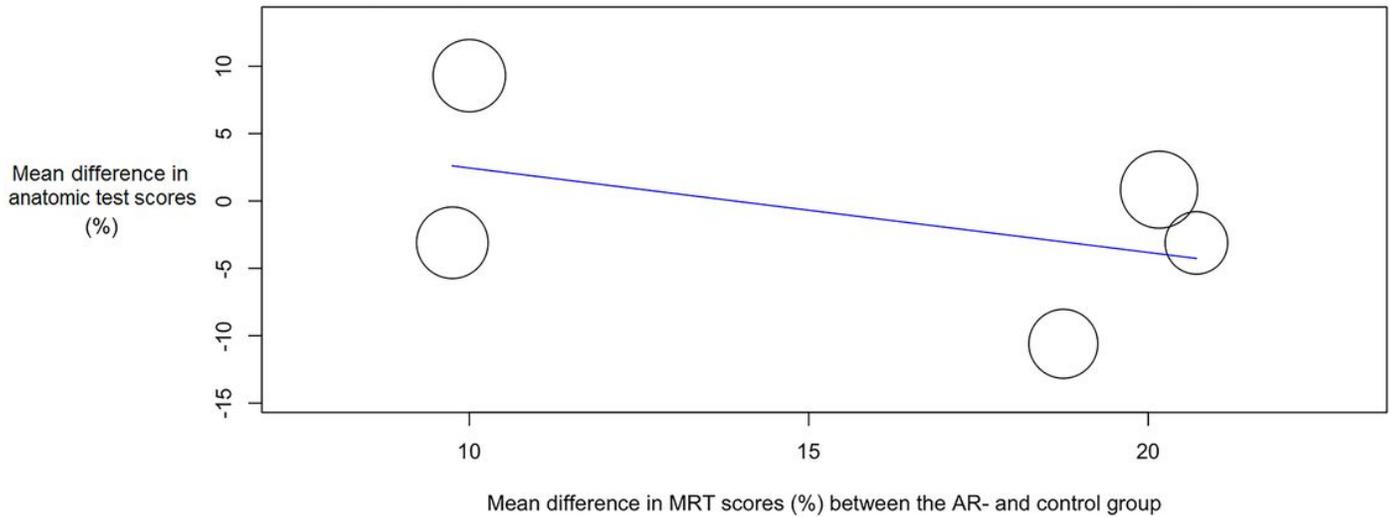


Figure 4
 Bubble plot with fitted meta-regression line of mean difference in anatomic test scores (%) and spatial ability Included are the studies addressing AR vs. 2D forms of anatomical education (i.e., traditional anatomical atlases, radiological data). AR= Augmented reality; MRT= Mental rotation test

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile.docx](#)