Supplementary Table. Details of the studies included in the review

	Author, Year, Country	AR-based tool used	Research Method	Data collection tool	Content Area	Study Participants	Study design (Evaluation of the AR tool)	Measured Outcome	Major Findings (Impact of the AR tool)
1	Henssen et al. 2020,[3] Netherlands	GreyMapp-AR	Mixed methods	Survey Focus group	Neuro-anatomy	1st year medical, biomedical students	Participants were randomly divided into two groups: Control and AR app Groups.	Knowledge acquisition Cognitive load Motivation Mental rotation	Students who worked with cross-sections in the control group (n = 16) showed significantly more improvement on test scores than students who worked with GreyMapp-AR (P = 0.035) (n = 15). No differences in - cognitive loads -MRT scores -motivation
2	Kugelmann et al. 2018[8] Germany	AR Magic Mirror	Quantitative	Survey	Anatomy	880 1st year medical students	Measured the effectiveness of an AR tool on student perception of learning	Engagement Spatial understanding	82% of participants agreed that AR facilitated engagement and active learning. 93% of participants agreed that AR improved their 3D understanding of human anatomy.
3	Ferrer-Torregrosa et al. 2015,[15] Spain	AR BOOK	Quantitative	Survey Test score	Anatomy	211 students from 7 public universities	Participants were randomly distributed into two groups: The control group received standard sessions. The experimental group received an AR book as well as the standard sessions.	Knowledge acquisition Metacognitive: Attention, motivation, autonomous learning, Spatial understanding	AR group showed higher scores for: -attention and motivation task. -autonomous work. -spatial comprehension. -written tests In the test the score (mean ± SD) for the Control group was 7.21 ± 1.73 points and 8.34 ± 1.64 points for the ARBOOK group. (p = 0.0001).
4	Bogomolova et al. 2020[18] Netherlands	Anatomical stereoscopic 3D AR model	Quantitative	Survey	Anatomy	1st and 2nd-year undergraduate students in medicine and biomedical sciences.	Compared between (i) stereoscopic 3D AR model, (ii) monoscopic 3D desktop model and (iii) 2D anatomical atlas. Randomized controlled trial.	Knowledge acquisition Mental rotation	- AR group performed as well on the knowledge test as the two other groups. The overall post-test scores in the stereoscopic 3D AR group (47.8%) were similar to those in the monoscopic 3D desktop group (38.5%; P = 0.240) and the 2D anatomical atlas group (50.9%; P = 1.00).

	Author, Year, Country	AR-based tool used	Research Method	Data collection tool	Content Area	Study Participants	Study design (Evaluation of the AR tool)	Measured Outcome	Major Findings (Impact of the AR tool)
									- Students in the AR group with lower MRT scores achieved higher post-test scores than those in the other two groups. Students with lower MRT scores achieved higher post-test scores in the stereoscopic 3D AR group (49.2%) as compared to the monoscopic 3D desktop group (33.4%; P = 0.015) and similar to the scores in the 2D group (46.4%; P = 0.99).
5	Bork et al. 2019[19] Germany	AR Magic Mirror	Quantitative	Survey Test score	Anatomy	749 1st year medical students	Students used the magic mirror (AR), anatomy table, or traditional radiology atlases. Pre and post-test scores were compared.	Knowledge acquisition Mental rotation	Significant improvements from pre to post-test scores (from 29.60 ± 18.37% to 64.89 ± 19.69% (P < 0.001). for the AR and anatomy table group. Students with low mental rotation test (MRT) scores benefited from the magic mirror (AR). They achieved significantly higher post-test scores than students with low MRT scores in the control group. For the MRT – High subgroup, the following average MRT scores resulted: Magic Mirror (91.54 ± 7.38%), Anatomage (87.07 ± 10.28%), and Theory (87.21 ± 9.66%). In the MRT – Low subgroup, the average MRT scores were 50.42 ± 10.48% for Magic Mirror, 52.55 ± 10.19% for Anatomage, and 51.92 ± 11.84% for the Atlas-based Theory group.
6	Küçük et al. 2016[20] Turkey	ARMagicBook	Mixed methods	Survey Interview	Neuro-anatomy	70 2nd year undergraduate students	Students were randomly distributed into experimental and control groups.	Knowledge acquisition Cognitive load	Students using AR tools: -scored higher on exams. The experimental group students who studied with ARMagicBook were significantly (P<0.05) more successful than the control group students.

	Author, Year, Country	AR-based tool used	Research Method	Data collection tool	Content Area	Study Participants	Study design (Evaluation of the AR tool)	Measured Outcome	Major Findings (Impact of the AR tool)
									-Moreover, the experimental group was found to have significantly (P<0.05) lower cognitive loads in comparison to the control group students.
7	Khan et al. 2019,[21] South Africa	Anatomy 4D	Quantitative	Survey	Anatomy	Undergraduate health science students studying medicine	Learning motivation was compared before and after the use of AR tool A questionnaire based on Keller's IMMS model for motivation was used to measure motivation.	Motivation Attention Confidence Satisfaction	-After using AR. the mean value significantly increased for attention (p<0.00001), confidence (p = 0.015) , and satisfaction (p = 0.0073), whiledecreased for relevance factor (p = 0.223) .
8	Ferrer-Torregrosa et al. 2016[22] Spain	ARBOOK	Quantitative	Survey Test score	Anatomy	170 health science students	Participants were divided into three groups: Didactic aid 1 group: Used supplied notes and traditional images. Didactic aid 2 group: Used supplied notes and video. Didactic aid 3 group: Used supplied notes and AR tool.	Knowledge acquisition Attention, Motivation, Autonomous learning, Spatial understanding	Compared to images and videos, students who used the AR tool had: i) Significantly higher test scores The average mark obtained with AR (7.20 points) is significantly higher than that obtained with video (6.54 points), which in turn is significantly higher than that obtained with the notes (5.61 points). ii) Higher scores in all aspects of metacognitive perceptions, including attention and motivation (p=0.001); autonomous learning (p=0.039); and Three-dimensional comprehension (p=0.004).
9	Norgaard et al. 2018,[23] Denmark	AR application for HoloLens	Mixed method	Survey Test score	Anatomy, CT-scan	110 University students	Participants were randomly distributed into three groups: Group 1: Used traditional PowerPoint	Knowledge acquisition Self-efficacy Motivation	-Quizzes helped the AR group gain spatial understanding. -No significant group differences in motivation test score. Self-efficacy was significantly higher for the group who used

	Author, Year, Country	AR-based tool used	Research Method	Data collection tool	Content Area	Study Participants	Study design (Evaluation of the AR tool)	Measured Outcome	Major Findings (Impact of the AR tool)
							Group 2: Used AR without quiz Group 3: Used AR with a quiz The goal was to examine the impact of AR app on students' self-efficacy and motivation, learning, and learning experience.	Spatial understanding	the AR tool along with quizzes ($p = 0.033$).
10	Gonzalez et al. 2020,[24] Chile	SPECTO	Quantitative	Survey Test score	Anatomy Physiology (of heart)	101 3rd year undergraduate biomedical science students.	Participants were randomly divided into two groups: the control group and the experimental group, who used AR. Pre- and post-test results were compared. Students were asked to do detailed anatomical drawings	Knowledge acquisition Motivation	AR use enhances the comprehension of anatomical and physiological concepts. Control and experimental groups showed no differences in baseline knowledge in their pre-test. The students who experienced the AR activities showed an increase in the complexity of representation levels in post-test results and also showed a significant difference in scores for the final exam. The use of AR increased motivation for learning
11	Schneider et al. 2020,[25] Australia	AR magic book	Mixed methods	Survey Test score	Pharmacy	25 undergraduate pharmacy students	Pre- and post-test scores Survey on student experiences using the tool Thematic analysis of written comments.	Knowledge acquisition Satisfaction	AR was effective to enhance student learning, showing 42% improvement in quiz score ($p < 0.0001$). -High acceptance -Engaging and stimulating
12	Duncan-Vaidya & Stevenson. 2020,[26] USA	Augmented Reality Head-Mounted Display	Quantitative	Survey	Anatomy	32 students from community college, studying	Group 1: Traditional books and models Group 2: HoloLens as the AR tool	Knowledge acquisition. Satisfaction	Pre and post-quiz scores were statistically the same between both groups. For anatomy group: pre-quiz = 32.7% (± 25.2); mean (± SD), post-quiz = 61.8% (± 19.5); $n = 15$; $t(28) = 3.53$; $P = 0.001$.

	Author, Year, Country	AR-based tool used	Research Method	Data collection tool	Content Area	Study Participants	Study design (Evaluation of the AR tool)	Measured Outcome	Major Findings (Impact of the AR tool)
						introductory anatomy and physiology courses.	The study compared: - pre- and post-quiz scores -engagement		For traditional group: pre-quiz = 44.9 % (± 18.6), post-quiz = 67.9 % (± 17.3); n = 17; t(32) = 3.73; P = 0.0007. The survey indicated: -AR was 'fun to use.' -AR was an effective and engaging tool for anatomy learning
13	Reeves et al. 2021,[27] UK	ZapWorks	Quantitative	Survey Test score	Structural Biology	20 University students	Participants were randomly assigned to one of the two groups: - Group 1 completed the formative test before attending the AR session. -Group 2 completed the AR session and then the formative test. Control group (Group 0): Did not attend any sessions (lecture or AR).	Knowledge acquisition Satisfaction	There was no statistically significant difference in test performance between groups 1 and 2 (those having completed the AR-session first versus the taking the quiz first), there was a significant difference in test performance between group 0 (no lectures and no AR session) and group 2, but not group 1. Students responded overwhelmingly positively to the engaging nature and interactivity of AR.
14	Noll et al. 2017,[28] Germany	mARble-Derma	Quantitative	Survey Test score	Dermatology	44 3rd-year medical students	Participants were randomly divided into two groups: A control group (B) and an experimental group (A)that used a mobile AR tool. Pre- and post-test scores were compared.	Knowledge acquisition	Pre and post-test improvements were similar between groups:group A: 3.59 [SD 1.48]; group B: 3.86 [SD 1.51]). Differences between both groups were statistically insignificant (P =.10) However, students who used the AR tool made 8.1% fewer errors on the test.

	Author, Year, Country	AR-based tool used	Research Method	Data collection tool	Content Area	Study Participants	Study design (Evaluation of the AR tool)	Measured Outcome	Major Findings (Impact of the AR tool)
15	ChanLin et al. 2019,[29] Taiwan	Mobile AR nutrition monitoring system	Mixed methods	Test score Interview	Nutrition	65 volunteer, non-nutrition major university students	Pre and post-tests were compared before and after the AR system use.	Knowledge acquisition	-Improvement in the mean nutritional concepts ($p < 0.01$) and a decrease in their mean misconceptions ($p < 0.001$) after learning with the Mobile AR nutrition monitoring system.
16	Albrecht et al. 2013,[30] Germany	mobile AR-based prototype app	Quantitative	Survey Test score	Gunshot wound	10 3rd year medical students.	Students were divided into control and AR-exposed groups. Pre and post-test score about gunshot wound was compared	Knowledge acquisition	The AR group (6/10) showed greater knowledge gain than the control group (4/10) (P =.03).
17	Herbert et al. 2021,[31] USA	Smartphone App on Heart Failure	Quantitative	Survey Test score (on Heart failure assessment)	Anatomy and physiology	33 Nursing students	A quasi-experimental, randomized pre-test post-test study was conducted. The experimental group used the self-paced app; the control group viewed the same content using pre-recorded video lecture.	Knowledge acquisition Satisfaction	No significant differences were found between the two groups for assessment completion time ($t(30)=1.626$, $p = .114$) and overall % test accuracy ($t(30)=1.846$, $p = .075$). No significant differences were found in "understanding" ($t(30)=1.058$, $p = .299$). Significant difference was reported between the students' accuracy on questions that required "remembering" ($t(30)=2.760$, $p = .010$). In the survey, 33.4% of students responded that they would prefer reviewing material via video lecture, 38.9% preferred the AR app
18	Mellos et al. 2022,[32] Australia	AR tool for the estimation food portions	Quantitative	Survey, Score (Assessment of the accuracy of the estimation of food portions)	Nutrition	33 University students studying nutrition.	A quasi-experimental, randomized pre-test post-test study was conducted. The experimental group Used the AR tool, while the control group	Knowledge acquisition	The AR group showed higher improvement between pre and post-test than the control group. The mean absolute error was lowest in the online group (53.0%), followed by AR (59.5%) and control (64.0%). Relative error scores revealed higher accuracy for the AR group (45.5%) followed by online

	Author, Year, Country	AR-based tool used	Research Method	Data collection tool	Content Area	Study Participants	Study design (Evaluation of the AR tool)	Measured Outcome	Major Findings (Impact of the AR tool)
							received an infographic instructing them how to estimate food portions with different hand shapes.		(43.5%), and control group (29.0%). Overall improvement in estimation was highest in the AR group (+12.2%) followed by the online (+11.6%) tool with a decrease seen for the infographic (−1.7%) tool.
19	Kim-Berman et al. 2019[33] USA	The AR virtual tooth identification test	Quantitative	Survey Test score	Dentistry, Tooth identification	93 first-year dental students	The AR virtual tooth identification test scores were compared with real tooth identification tests, scores on summative exams. A survey was also conducted.	Knowledge acquisition Validity of the AR tool.	The AR virtual tooth identification test had a positive correlation with the real tooth identification test ($r=0.410$, $p<0.01$), a combined score of two real tooth identification tests ($r=0.545$, $p<0.01$), the final exam ($r=0.489$, $p<0.01$), and overall grade for the dental anatomy course ($r=0.661$, $p<0.01$). The students had some difficulty viewing images and experienced technical difficulties related to their smartphones, and their survey responses expressed little support for the AR tool.