

Seeing Disease in Three Dimensions: A Comparison of Extended Reality and Traditional Lectures in Pre-Medical Pathology Education

Iris S. Wang

iriswsr@umich.edu

Mixel Studio, Inc

Yu-Chiao Lin

Mixel Studio, Inc

Swetha Konduru

University of Michigan

Camille van Neste

Michigan Medicine

Caleb Vogt

Michigan Medicine

Anhua Wu

Mixel Studio, Inc

Kamran M. Mirza

Michigan Medicine

Research Article

Keywords: Extended reality, virtual reality, pathology education, pre-medical education, spatial learning, knowledge retention

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10944095/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. Three of the seven authors (I.W., Y-C.L., and A.W.) are affiliated with Mixel Studio, Inc., the company that develops the OXR software evaluated in this study.

This constitutes a potential competing interest, which we have described in full in the Declarations section of the manuscript.

Abstract

Background

Extended reality (XR) holds potential for pre-medical pathology education by visualizing complex three-dimensional spatial relationships, though hardware constraints often limit adoption. This study investigated the efficacy of video lectures using OXR, a scalable, web-based XR tool compared to traditional two-dimensional (2D) video lectures.

Methods

In a format-counterbalanced crossover study, 28 undergraduate students learned about disease progression at epithelial transition zones (Barrett's esophagus and HPV-driven cervical cancer).

Results

Students overwhelmingly preferred the OXR format, rating it significantly higher for spatial understanding (OXR: 4.25 vs. 2D: 3.46, $p = 0.017$), engagement ($p = 0.001$), and overall preference ($p = 0.011$). Despite these subjective preferences, immediate knowledge acquisition showed no significant difference between formats in either module, though scores trended lower for OXR in the esophagus module (OXR: 3.62 vs. 2D: 4.33, $p = 0.070$) but not in the cervix module ($p = .682$). Delayed retention tests administered 7–10 days later demonstrated robust overall learning gains (+ 4.04 points, $p < 0.001$) but no statistical difference between the two delivery formats.

Conclusions

While students strongly preferred OXR and rated it significantly better for conveying spatial relationships, this did not translate into improved knowledge scores or retention compared to 2D. Scalable, web-based XR tools such as OXR appear promising for pre-medical education, though how best to integrate them remains to be studied.

Background

Extended reality (XR), which encompasses virtual reality (VR), augmented reality (AR), and mixed reality (MR), is increasingly recognized as a transformative technology in medical and pre-medical education [1]. Traditionally, foundational pre-medical education has relied on textbooks, didactic lectures, and two-dimensional (2D) images, which often fall short in adequately conveying complex spatial relationships and dynamic biological processes [2,3]. By offering immersive, interactive, and risk-free simulated environments, XR allows learners to visualize and manipulate three-dimensional (3D) anatomical and pathological structures in real-time [4]. Furthermore, integrating XR into curricula mitigates the financial,

ethical, and supervisory constraints associated with traditional educational tools, such as cadaveric dissection and specialized laboratory equipment [5,6].

In anatomical education, XR technologies have demonstrated significant potential to transcend the limitations of traditional 2D media by providing stereoscopic, manipulable 3D models that enhance spatial understanding and lower cognitive load [2,3]. A comprehensive meta-analysis [3] revealed that XR technologies yield higher knowledge gains than traditional didactic lectures, textbooks, and atlases, particularly when utilized as a complementary learning resource rather than a complete replacement. However, the evidence on objective academic performance remains nuanced. Other meta-analyses indicate that while XR enhances the educational experience, it does not always produce statistically superior test scores compared to physical models or 2D learning [7]. Despite these mixed academic outcomes, studies consistently show that students overwhelmingly prefer immersive formats, citing increased engagement, motivation, and perceived usefulness for grasping complex spatial relationships [8–10].

Unlike pure anatomy, pathology education adds a further layer of difficulty, since it requires linking microscopic histological features with macroscopic clinical and radiographic manifestations, a multi-dimensional integrative process that can be challenging for novices to conceptualize [11]. While digital pathology and whole-slide imaging have successfully modernized microscopic learning [12], there remains a critical gap in connecting 2D histological slides to the 3D structural and anatomical changes caused by disease. Immersive tools like VR can contextualize pathological processes, helping students build mental schemas that connect cellular abnormalities to gross anatomical consequences [11,13]. Recent innovations, such as interactive 360° VR tours of clinical pathology laboratories, have been introduced to immerse students in clinical environments, fostering high interactivity and engagement during remote learning periods [14]. Additionally, pioneering mixed reality systems that integrate anatomy, pathology, and physical examination into a unified experience have shown that contextualizing disease states within a 3D anatomical framework deepens holistic clinical understanding [13]. Nevertheless, empirical studies measuring the objective learning outcomes of XR interventions specifically tailored for theoretical pathology remain scarce, presenting a critical gap in the current literature [15].

While XR offers immense pedagogical potential, its widespread implementation in medical education faces significant scalability and accessibility challenges. Most XR interventions rely heavily on head-mounted displays (HMDs), which are associated with high procurement costs, complex technical setups, and ongoing maintenance requirements [4,6]. Providing individual HMDs for large cohorts of undergraduate pre-medical students is often financially and logistically unfeasible for institutions, creating inequities in access to these advanced learning tools [2,16]. Furthermore, HMDs frequently induce adverse physiological side effects in users, such as cybersickness, nausea, headaches, and blurred vision, which can prematurely interrupt the learning experience [9,17]. To circumvent these hardware and physiological barriers, educators are increasingly exploring virtual production and live-streaming XR setups. Systems like XRStudio enable instructors to manipulate immersive 3D

environments in VR while broadcasting the experience as a live or recorded 2D video feed to students on standard screens [18]. This instructor-led broadcasting approach drastically reduces hardware dependencies, making the spatial and dynamic benefits of XR accessible at scale without requiring students to wear HMDs.

To address the intersection of these educational gaps, the present study utilizes OXR, a scalable, web-based XR tool that empowers educators to manipulate 3D pathology and anatomy objects in immersive environments while students participate via live web streaming. OXR is best understood as a system of spatial slides. Like a conventional slide deck, OXR organizes instruction into discrete pages, or scenes. Each scene loads a set of pre-authored anatomical materials, digital three-dimensional organ models, images, and information cards, which the instructor presents through a third-person virtual camera (Figure 1). For the purposes of this controlled comparison, both lecture formats were pre-recorded as video lectures using identical narration, rather than delivered live, in order to hold delivery timing and audio constant across conditions.

By utilizing this learner-side HMD-free delivery method, this research bypasses the traditional hardware and physiological constraints of VR, providing equitable and scalable access to spatial learning. Junior and senior undergraduate health sciences students from the University of Michigan were recruited to participate in a format-counterbalanced crossover study comparing the OXR 3D slide-based lecture format against traditional 2D slide-based lecture format. The curriculum focused on disease progression, tracing the narrative from normal epithelium through metaplasia, dysplasia, neoplasia, and metastasis, applied specifically to Barrett's esophagus and HPV-driven cervical cancer. By assessing both immediate knowledge gains and delayed retention, alongside subjective perceptions of effectiveness, this study aims to delineate the true value of scalable, instructor-led XR video lectures in pre-medical pathology education.

Methods

Participants

Participants were junior and senior undergraduate students enrolled in health sciences programs at the University of Michigan. All participants provided informed consent prior to participation and were informed that participation was voluntary and that they could withdraw at any time without penalty. A target sample of 52 participants was determined a priori via G*Power based on a medium effect size ($d = 0.40$) drawn from a meta-analysis of XR in anatomy education [3], using a two-tailed paired-samples design at $\alpha = .05$ and power = .80. This was the more conservative of two effect sizes reported in that source (knowledge gain: $SMD = 0.40$; perceived usefulness: $SMD = 0.54$). We powered on the smaller estimate to ensure adequate sensitivity across both the primary perception outcomes and the secondary knowledge-gain outcomes. The final sample consisted of 28 participants (15 assigned to the 2D → OXR sequence; 13 assigned to the OXR → 2D sequence). Analyses are interpreted with the recognition that

this sample is underpowered relative to the target, and effect sizes are therefore reported and emphasized alongside p-values.

Stimuli

The lecture content covered epithelial transition zones (junctional zones), with two anatomical sub-topics presented within a single session: the gastroesophageal (GE) junction (Topic A) and the endo-ecto cervical junction (Topic B). Both topics introduced the concept of squamo-columnar junctions, explained the epithelial lining of each region, and followed the same disease progression narrative from normal epithelium through metaplasia, dysplasia, neoplasia, and metastasis, applied respectively to Barrett's esophagus (GE junction) and HPV-driven cervical cancer (cervical junction).

For each topic, two instructional videos were produced: one in the control two-dimensional slide format (2D condition) and one in the experimental extended reality format (OXR condition), yielding four videos in total. To isolate the effect of visual modality, all other production elements were held constant across conditions. Both videos for a given topic used the identical pre-recorded audio track with the same script, narration, and timing, ensuring that any differences in learning outcomes could not be attributed to variation in verbal content, pacing, or instructor delivery. Video duration was matched across conditions within each topic, and all videos were delivered embedded in Qualtrics with limited playback controls to minimize differences in viewing behavior.

Design

The study used a within-subject design. Each participant viewed both the OXR and 2D video formats and both content topics, allowing direct paired comparisons of format effects within individuals while controlling for individual differences in prior knowledge, spatial ability, and general engagement. Topic order was fixed for all participants: the esophagus topic was always presented first (Part 1) and the cervix topic was always presented second (Part 2).

Two between-subject factors were counterbalanced to control for format-order and content-assignment effects: Sequence (OXR → 2D vs. 2D → OXR) and Content Assignment (Topic A in OXR / Topic B in 2D vs. Topic A in 2D / Topic B in OXR). Participants were randomized 1:1 by the study platform to one of these counterbalancing groups. This design permits separation of format effects (OXR vs. 2D) from content effects (Topic A vs. Topic B) and sequence effects (first viewed vs. second viewed).

The primary outcomes were student perceptions of the two lecture formats, assessed in two ways: an immediate post-video perception survey completed right after each lecture format, measuring overall effectiveness, ability to convey spatial relationships, engagement and motivation, and confidence in the material; and an end-of-study comparative study completed after both formats had been viewed, directly rating the two formats against one another. Secondary outcomes were immediate knowledge gain (post-

test score adjusted for pre-test baseline) and delayed retention (assessed 7–10 days after the study session).

Procedure

The study was conducted in a quiet, proctored university classroom equipped with stable Wi-Fi and power access. Participants brought their own laptops; over-ear headphones were available on loan. The session proceeded as follows.

Participants first completed a pre-test consisting of 10 multiple-choice questions with five aligned to Topic A and five aligned to Topic B, drawn from a question bank of 20 items (10 per topic). The initial item pool was generated using ChatGPT (OpenAI) and subsequently reviewed, revised, and validated for content accuracy by a board-certified pathologist (K.M.M.) prior to use. This baseline assessment established prior knowledge for both topics prior to any instructional exposure and was used to compute knowledge gain scores.

Participants were then randomized to their counterbalancing condition by the study platform and proceeded through the two instructional blocks in their assigned sequence. For each block, participants first viewed their assigned video lecture for that topic (approximately 5 minutes). Immediately following the video, they completed a post-test consisting of five multiple-choice questions aligned to that topic (approximately 2 minutes), drawn as a parallel form from the same question bank as the pre-test. Participants then completed a brief perceptions survey for that lecture format (approximately 2 minutes), consisting of five items assessing overall effectiveness, ability to convey spatial relationships, engagement and motivation, and confidence. This procedure was then repeated for the second topic in the opposite format condition.

After completing both instructional blocks, participants completed an end-of-study survey consisting of seven Likert-scale items and five open-ended questions. Scale items included direct comparative ratings of the two formats (e.g., "Overall, which format did you find more effective for learning?") anchored on a 5-point scale centered at 3 ("no difference"). Open-ended items invited participants to elaborate on format preferences, perceived strengths and limitations of each approach, and suggestions for improvement. The total in-person session lasted approximately 25–30 minutes.

A delayed retention test was administered online 7–10 days after the in-person session. This test consisted of 10 multiple-choice questions with five per topic and drawn randomly from the question bank as a new parallel form, distinct from both the pre-test and the post-tests. Participants completed this assessment independently and were reminded of the test by a follow-up email at the 7-day mark.

Data Analysis

All primary analyses used non-parametric tests appropriate for ordinal and potentially non-normally distributed data, given the scale and integer-valued nature of the outcome measures. Parametric

equivalents were reported alongside non-parametric results where applicable, for comparability with prior literature.

Knowledge gain was operationalized as the post-test score minus the corresponding pre-test subscale score for each topic. The effect of format (OXR vs. 2D) on knowledge gain was assessed using a Wilcoxon Signed-Rank Test, which compared each participant's OXR gain score against their 2D gain score as a within-subject paired comparison. A paired t-test was also conducted and reported alongside the Wilcoxon test. Delayed retention scores were analyzed using the same approach, comparing OXR and 2D retention scores within participants.

To examine potential content and order effects, Mann-Whitney U Tests were used for independent group comparisons: specifically, to compare knowledge gain scores between participants who received Topic A in the OXR condition versus those who received Topic A in the 2D condition (and equivalently for Topic B), and to test whether the magnitude of the OXR–2D difference score varied by sequence (OXR-first vs. 2D-first).

Perception survey items from the post-video surveys were analyzed using Wilcoxon Signed-Rank Tests to compare OXR and 2D ratings within participants for each dimension (overall effectiveness, spatial relationships, engagement/motivation, confidence). End-of-study comparative items (rated on a 5-point scale centered at 3) were tested against the scale midpoint using one-sample Wilcoxon Signed-Rank Tests, which assessed whether each item was systematically above or below 3 (indicating a directional preference for one format).

Effect sizes and 95% confidence intervals were calculated for all tests and are reported throughout, given that the final sample of 28 is below the a priori target of 52. For independent-samples comparisons (knowledge acquisition by topic, retention by topic, and order $\times$ condition interactions), rank-biserial correlation r [19] is reported alongside the primary Mann-Whitney U test and interpreted qualitatively ($r < .10$ trivial, $.10$ – $.30$ small, $.30$ – $.50$ medium, $> .50$ large). For paired-samples and one-sample comparisons (perception surveys), Cohen's d is reported instead: for several items, r was internally inconsistent with the test's own p -value (e.g., a significant result with $r \approx 0$), likely reflecting instability of the tie-sensitive rank-biserial formula in this ordinal, small-sample data. Non-significant p -values are not interpreted as evidence of no effect given the underpowered sample.

Results

Participants and Baseline

Twenty-eight undergraduate students completed the study across three sessions (Session 1: $n = 10$; Session 2: $n = 13$; Session 3: $n = 5$). Thirteen participants were randomized to view the OXR lecture first (OXR-first group), and 15 viewed the 2D lecture first (2D-first group). Twenty-seven of the 28 participants completed the delayed retention test.

Pre-assessment quiz scores were low across the full sample ($M = 2.18$, $SD = 1.22$, $Mdn = 2.00$, range: 0–4 out of 10), confirming that participants had minimal prior knowledge of the content before the intervention. The two order groups did not differ significantly on pre-assessment performance (OXR-first vs. 2D-first: $t(26) = 0.21$, $p = .837$, Cohen's $d = 0.08$, 95% CI [-0.665, 0.821]; Mann-Whitney $U = 97.5$, $p = 1.000$, rank-biserial $r = 0.00$), indicating that randomization produced comparable groups at baseline.

Shapiro-Wilk testing confirmed significant departures from normality for several outcome variables, consistent with the non-parametric approach specified above.

Immediate Perception Surveys

Following each lecture, participants rated that format on five items using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) (Table 1).

Table 1
Immediate Perception Ratings by Condition

Perception Item	OXR M (SD)	2D M (SD)	W	p	t(27)	p	d [95% CI]
Effective	3.89 (1.13)	3.96 (1.00)	33.0	.654	-0.40	.691	-0.08 [-0.446, 0.296]
Spatial relationships	4.25 (1.27)	3.46 (1.35)	126.5	.017	2.53	.017	0.48 [0.083, 0.866]
Engagement	3.89 (1.17)	3.68 (1.19)	75.0	.397	1.00	.326	0.19 [-0.187, 0.561]
Confidence	3.89 (1.07)	3.71 (0.98)	74.0	.431	0.76	.456	0.14 [-0.231, 0.514]
Recommend	3.75 (1.21)	3.79 (1.13)	27.0	1.000	-0.21	.832	-0.04 [-0.411, 0.330]
Overall mean	3.94 (0.99)	3.72 (0.88)	211.0	.083	1.34	.191	0.25 [-0.125, 0.628]
Note. W = Wilcoxon signed-rank statistic (primary test); t(27) and d (Cohen's d) = paired-samples t-test and effect size, reported for comparability with prior literature.							

Of the five perception items, only spatial relationships reached statistical significance. Participants rated the OXR lecture as significantly better at conveying spatial relationships compared to the 2D lecture (Wilcoxon $W = 126.5$, $p = .017$, Cohen's $d = 0.48$, 95% CI [0.083, 0.866]).

End-of-Study Comparative Survey

At the end of the session, participants completed a seven-item comparative survey rating the two formats on a 5-point scale (1 = strongly favors 2D, 3 = no difference, 5 = strongly favors OXR). One-

sample tests assessed whether each item differed significantly from the scale midpoint of 3 (Table 2).

Table 2
End-of-Study Comparative Survey Results (N = 28; Test Value = 3.0)

Item	M (SD)	W	p	t(27)	p	d [95% CI]
Understanding content	3.21 (1.03)	111.0	.248	1.10	.281	0.21 [-0.169, 0.580]
Spatial relationships	4.46 (0.64)	351.0	< .001	12.16	< .001	2.30 [1.579, 3.005]
Engagement / motivation	3.75 (0.97)	189.0	.001	4.10	< .001	0.78 [0.346, 1.194]
Confidence	3.07 (0.86)	59.5	.660	0.44	.663	0.08 [-0.289, 0.454]
Ease of following	3.46 (1.00)	136.0	.022	2.46	.021	0.46 [0.070, 0.851]
Quiz preparation	2.96 (1.00)	81.0	.852	-0.19	.851	-0.04 [-0.406, 0.335]
Overall preference	3.57 (1.03)	217.0	.011	2.93	.007	0.55 [0.150, 0.947]
Composite mean	3.50 (0.72)	317.0	.002	3.67	.001	0.69 [0.275, 1.102]
Note. W = one-sample Wilcoxon signed-rank statistic (primary test); t(27) = one-sample t-test, reported for comparability.						

Four of the seven items differed significantly from the midpoint, all in the direction favoring OXR. The spatial relationships item produced by far the largest effect ($d = 2.30$, 95% CI [1.58, 3.01]), reflecting both a substantial mean shift and low response variability ($SD = 0.64$). Engagement/motivation, ease of following, and overall preference also significantly favored OXR (all $p < .05$, $d = 0.46$ – 0.78). The composite mean across all seven items was significantly above the midpoint (Wilcoxon $W = 317.0$, $p = .002$, $d = 0.69$, 95% CI [0.28, 1.10]), indicating an overall comparative preference favoring the OXR format.

Immediate Knowledge Acquisition

Overall Paired Comparison

Each participant's quiz score from the OXR lecture was paired with their score from the 2D lecture for a within-subject comparison. Participants scored somewhat lower in the OXR condition ($M = 3.68$, $SD = 1.06$) than the 2D condition ($M = 4.11$, $SD = 0.875$), a difference that approached but did not reach conventional significance (Wilcoxon $W = 79.5$, $p = .065$, rank-biserial $r = -0.61$; paired $t(27) = -1.89$, $p = .069$, Cohen's $d = -0.36$, 95% CI [-0.736, 0.028]).

Lecture Part 1 (Esophagus) vs. Lecture Part 2 (Cervix)

Because all participants viewed the esophagus topic first and the cervix topic second, topic-level comparisons were conducted as between-group analyses, comparing the subgroup who received OXR in that position against those who received 2D.

For the esophagus lecture (Part 1), the 2D group scored notably higher than the OXR group (2D: $M = 4.33$, $SD = 0.488$, $n = 15$; OXR: $M = 3.62$, $SD = 1.12$, $n = 13$), though on the primary (Mann-Whitney U) test this difference was only a non-significant trend. Levene's test indicated a significant violation of the equal-variances assumption for this comparison, so Welch's t-test was used in place of Student's t-test. Both the Welch-corrected and uncorrected Student's tests reached significance and are reported here for comparability with prior literature (see Table 3 for full test statistics).

For the cervix lecture (Part 2), no meaningful difference between conditions was observed (OXR: $M = 3.73$, $SD = 1.03$, $n = 15$; 2D: $M = 3.85$, $SD = 1.14$, $n = 13$; see Table 3 for full test statistics).

Table 3
Immediate Quiz Scores by Topic and Condition (Out of 5)

Topic	Condition	n	M (SD)	U / t	p	r / d [95% CI]
Esophagus (Part 1)	OXR	13	3.62 (1.12)	U = 60.0	.070	$r = 0.385$
	2D	15	4.33 (0.49)			
	Student's t			$t = -2.25$	.033	$d = -0.85$ [-1.623, -0.068]
	Welch's t			$t = -2.14$ (df = 15.9)	.048	$d = -0.83$
Cervix (Part 2)	OXR	15	3.73 (1.03)	U = 88.5	.682	$r = -0.092$
	2D	13	3.85 (1.14)			
	Student's t			$t = 0.27$	.786	$d = 0.10$ [-0.640, 0.846]

Note. U = Mann-Whitney U Statistic (primary test); r = rank-biserial correlation. Welch's t reported in place of Student's t where Levene's test indicated a significant violation of the equal-variances assumption; d = Cohen's d. Group sizes differ by topic because order assignment determined which condition was viewed first.

Delayed Knowledge Retention

Overall Learning Gain

Retention test scores (out of 10) were available for 27 of the 28 participants ($M = 6.19$, $SD = 2.17$, $Mdn = 6.0$). In this same matched sample, pre-assessment scores averaged $M = 2.15$ ($SD = 1.23$, $Mdn = 2.00$).

Retention scores were substantially and significantly higher than pre-assessment scores, representing a robust mean learning gain of 4.04 points out of 10 (paired $t(26) = 8.65$, $p < .001$, Cohen's $d = 1.66$, 95% CI [1.07, 2.24]; Wilcoxon $W = 376$, $p < .001$, rank-biserial $r = 0.987$). This very large effect size confirms that the intervention produced substantial and statistically robust learning regardless of condition.

Retention by Condition

Because the delayed retention test drew questions randomly from a 20-item bank (10 per topic), the number of questions each participant received per topic varied. Retention was therefore expressed as proportion correct per topic to enable valid comparisons across participants (Table 4). Across both topics, condition-level differences in long-term retention were non-significant, suggesting that the OXR and 2D formats produced broadly equivalent long-term knowledge retention.

Table 4
Delayed Retention by Topic and Condition (Proportion Correct, $n = 27$)

Topic	Condition	M (SD)	U	p	r
Esophagus	OXR	0.69 (0.25)	83.5	.727	0.082
	2D	0.63 (0.28)			
Cervix	OXR	0.56 (0.28)	67.5	.250	-0.258
	2D	0.62 (0.29)			
Note. U = Mann-Whitney U statistic (primary test); r = rank-biserial correlation. Proportion correct used because the number of retention items drawn per topic varied across participants (range: 3–7 items per topic).					

Order Effects and Condition × Sequence Interactions

Given the format-counterbalanced design, we examined whether the difference between OXR and 2D scores varied as a function of which format participants encountered first. For each participant and outcome, a difference score (OXR minus 2D) was computed, and these difference scores were compared between the OXR-first and 2D-first groups.

No significant order interaction was found for quiz scores and most perception items. However, two perception items showed significant or marginally significant crossover interactions. For confidence, a significant interaction emerged consistently across all three tests, confirming a robust crossover effect. For recommend, Levene's test indicated a significant violation of the equal-variances assumption; the Welch-corrected test was no longer significant, in contrast to the uncorrected Student's result, and the Mann-Whitney U test likewise showed only a non-significant trend (see Table 5 for full test statistics).

Table 5
Condition x Order Interaction: Difference Scores (3D – 2D) by Order Group

Outcome	OXR-first M	2D-first M	U	p	r	t	d
Quiz score	-0.23	-0.60	81.0	.446	-0.169	0.81	0.31 [-0.444, 1.050]
Effective	-0.15	0.00	87.0	.608	0.108	-0.43	-0.16 [-0.904, 0.584]
Spatial	+ 0.85	+ 0.73	88.5	.683	-0.092	0.18	0.07 [-0.676, 0.810]
Engagement	-0.08	+ 0.47	75.5	.293	0.226	-1.28	-0.49 [-1.235, 0.274]
Confidence	-0.46	+ 0.73	47.5	.016	0.513	-2.84	-1.07 [-1.863, -0.268]
Recommend	-0.39	+ 0.27	64.5	.079	0.339	-2.07	-0.78 [-1.548, -0.003]
Note. M values represent mean within-subject difference scores (OXR minus 2D) for each order group. df = 26 for Student's t-tests. U = Mann-Whitney U statistic (primary test); r = rank-biserial correlation. Confidence reached significance ($p < .05$) on the primary test.							

These interaction patterns indicate that first exposure to the OXR format, without prior contextual experience of 2D content, was associated with significantly lower perceived confidence in the OXR condition. Participants who viewed 2D first and then OXR rated the OXR format more favorably on confidence specifically, a pattern consistent with a contrast effect, in which the 2D baseline experience enhanced the perceived value of the subsequent OXR experience. Critically, the spatial understanding advantage of OXR was unaffected by presentation order: both order groups rated OXR as superior for conveying spatial relationships, with no significant difference in the magnitude of that advantage between groups ($p = .860-.865$ across all three tests), supporting the robustness and order-independence of this finding.

Discussion

Students in this study strongly preferred the OXR format, rating it significantly higher for spatial understanding, engagement, and overall preference in the end-of-study comparison, consistent with existing literature showing that students favor extended reality (XR) formats for conveying spatial relationships and fostering engagement [9,21]. Yet this enthusiasm did not translate into a corresponding advantage in objective performance: immediate knowledge scores and delayed retention were statistically equivalent between formats, and where a difference trended in either direction, it favored 2D rather than OXR.

Several factors may account for this dissociation, though our design does not allow us to test them directly. One possibility is a "novelty effect" combined with the additional cognitive load of orienting to an unfamiliar XR interface, which could draw attention away from core content during initial exposure [9,20]; cognitive load was not measured directly, and this is advanced only as a hypothesis for future work. A second possibility is methodological: the exclusive use of multiple-choice questions (MCQs) may lack the sensitivity required to capture the specific 3D spatial knowledge that XR is best positioned to convey [5,21]. If the perceptual spatial advantage we observed does correspond to genuine differences in spatial understanding, assessments designed to probe spatial reasoning directly, rather than factual recall, may be needed to detect it.

An order effect for learner confidence was also observed: participants who encountered OXR first, without a prior 2D foundation, reported lower confidence in that format than those who saw 2D first. This suggests that first-time exposure to OXR may shape learners' confidence specifically; whether this extends to a broader principle about how XR should be sequenced within a curriculum is a reasonable hypothesis but remains untested here, since this order effect was not observed for quiz performance or for the spatial-relationship finding, both of which were stable across presentation order.

A central feature of the present design is that it allows the observed spatial-relationship advantage to be interpreted more precisely than in much prior XR research. The OXR and 2D conditions were matched on narration, content, pacing, and a shared page-based, pre-authored instructional structure. As a result, the primary instructional contrast was whether the same material was presented as a third-person 3D spatial scene or as a 2D plane. Notably, OXR produced this spatial-perception benefit without relying on an instructor avatar or extensive live free-hand sketching as the primary means of content generation. Instead, the instructional content was pre-authored, while real-time interaction was limited mainly to deictic guidance, such as pointing, circling, and annotating key structures within the spatial scene.

This distinction is important because the spatial-relationship advantage was both the largest perceptual effect observed in the study and robust to presentation order. Both order groups rated OXR as superior for conveying spatial relationships, by a similar margin. This suggests that the perceived spatial benefit may be attributable less to instructor embodiment or open-ended immersive performance, and more to the combination of presentational dimensionality, pre-authored spatial content, and lightweight deictic guidance. At the same time, this perceptual advantage did not translate into higher immediate quiz scores or delayed retention. Thus, in this sample and with the assessment measures used here, spatial dimensionality appeared to shape how clearly learners perceived anatomical and pathological relationships, but not what they measurably retained.

Limitations

This study has several limitations. First, and most notably, the final sample of 28 participants fell well short of the a priori target of 52 determined by our power analysis. Recruitment was constrained to a single academic semester, and scheduling each in-person session required coordinating mutual

availability across the company lead, the supervising professor, and the primary researcher, which limited data collection to three sessions. Despite recruitment efforts spanning flyers, student clubs, and campus associations that enrolled more than 52 students, only 28 ultimately attended an in-person session and completed the study. As a result, the study is underpowered relative to its planned design, and non-significant findings should not be interpreted as evidence of no effect; effect sizes are reported throughout to support interpretation independent of statistical power.

Second, immediate knowledge acquisition was assessed exclusively using MCQs, which may lack the sensitivity required to capture the specific 3D spatial understanding that XR is best positioned to convey; assessments designed to probe spatial reasoning directly could reveal effects that MCQs do not. Third, participants were drawn from a single institution and a specific pre-medical health sciences population, which may limit generalizability to other learner populations or educational contexts. Finally, because all participants viewed the esophagus topic before the cervix topic, topic-order and format-order were not fully disentangled, and the study cannot rule out some residual influence of topic sequence on the pattern of results.

Conclusions

This study demonstrates the feasibility and strong student preference for scalable, learner-HMD-free XR delivery methods in pre-medical pathology education. By utilizing a web-based approach, educators can bypass the high costs, logistical challenges, and adverse physiological effects typically associated with immersive virtual reality, promoting more equitable access to advanced spatial learning tools.

Students strongly preferred the OXR format and rated it significantly better at conveying spatial relationships, yet showed no significant difference in immediate knowledge scores or delayed retention compared to 2D. This dissociation between subjective experience and objective performance may reflect the added cognitive load of a novel interface. The strength of student preference and perceived spatial understanding suggests scalable, web-based XR tools like OXR represent a promising direction for pre-medical curricula; how such tools are best integrated, and whether this perceptual advantage can translate into measurable learning gains, remains to be studied.

Abbreviations

2D: Two-dimensional **3D:** Three-dimensional **AR:** Augmented reality **GE:** Gastroesophageal **HMD:** Head-mounted display **HPV:** Human papillomavirus **MCQ:** Multiple-choice question **MR:** Mixed reality **OXR:** Web-based Extended Reality Tool **VR:** Virtual reality **XR:** Extended reality

Declarations

Ethic approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. This study was reviewed and approved by the University of Michigan Institutional Review Board (IRB# HUM00283735). All participants were provided informed consent prior to participation.

Clinical trial number

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

Three of the seven authors (I.W., Y-C.L., and A.W.) are employees of Mixel Studio, Inc., the company that develops the OXR software evaluated in this study. This constitutes a potential competing interest, as a favorable evaluation of OXR could benefit the company commercially. S.K. was a student at the University of Michigan at the time of this study and has no affiliation with Mixel Studio, Inc. K.M.M., who conceived and supervised the study and provided pathology content expertise, is affiliated with Michigan Medicine, University of Michigan, and has no financial or other competing interest in Mixel Studio, Inc. or OXR. I.W. held the data and conducted the formal analysis.

Funding

Participant gift cards were provided from K.M.M.'s educational funds at Michigan Medicine, University of Michigan. The study did not receive any external grant funding from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

IW led study design, stimulus development, formal analysis, and manuscript drafting, and contributed to data collection. Y-CL led data collection and pathology lecture design and development, and contributed to writing the sections explaining OXR functionality. SK led pathology lecture design. CVN and CV contributed to the IRB application. AW contributed to OXR product development. KMM conceived and supervised the study, provided pathology content expertise, and provided feedback on the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Curran VR, Xu X, Aydin MY, Meruvia-Pastor O. Use of Extended Reality in Medical Education: An Integrative Review. *Med Sci Educ.* 2023;33:275–86. <https://doi.org/10.1007/s40670-022-01698-4>
2. Barrow J. The Future with Extended Reality, Three-Dimensional, and Advanced Imaging for Molecules, Microscopy, and Anatomy. In: Witchel HJ, Lee MW, editors. *Technol Biomed Life Sci Educ Approaches Evid Effic Learn* [Internet]. Cham: Springer International Publishing; 2022 [cited 2026 June 28]. p. 519–57. https://doi.org/10.1007/978-3-030-95633-2_17
3. García-Robles P, Cortés-Pérez I, Nieto-Escámez FA, García-López H, Obrero-Gaitán E, Osuna-Pérez MC. Immersive virtual reality and augmented reality in anatomy education: A systematic review and meta-analysis. *Anat Sci Educ.* 2024;17:514–28. <https://doi.org/10.1002/ase.2397>
4. Barteit S, Lanfermann L, Bärnighausen T, Neuhaus F, Beiersmann C. Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. *JMIR Serious Games.* JMIR Publications Inc., Toronto, Canada; 2021;9:e29080. <https://doi.org/10.2196/29080>
5. Gnanasegaram J, Leung R, Beyea J. Evaluating the effectiveness of learning ear anatomy using holographic models. *J Otolaryngol - Head Neck Surg.* 2020;49. <https://doi.org/10.1186/s40463-020-00458-x>
6. Herur-Raman A, Almeida ND, Greenleaf W, Williams D, Karshenas A, Sherman JH. Next-Generation Simulation—Integrating Extended Reality Technology Into Medical Education. *Front Virtual Real* [Internet]. Frontiers; 2021 [cited 2026 June 28];2. <https://doi.org/10.3389/frvir.2021.693399>
7. Moro C, Birt J, Stromberga Z, Phelps C, Clark J, Glasziou P, 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.* 2021;14:368–76. <https://doi.org/10.1002/ase.2049>
8. Al-Hor M, Almahdi H, Al-Theyab M, Mustafa AG, Seed Ahmed M, Zaqout S. Exploring student perceptions on virtual reality in anatomy education: insights on enjoyment, effectiveness, and preferences. *BMC Med Educ.* 2024;24:1405. <https://doi.org/10.1186/s12909-024-06370-6>
9. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anat Sci Educ.* 2017;10:549–59. <https://doi.org/10.1002/ase.1696>
10. Nakamatsu NA, Aytaç G, Mikami B, Thompson JD, Davis M, Rettenmeier C, et al. Case-based radiological anatomy instruction using cadaveric MRI imaging and delivered with extended reality web technology. *Eur J Radiol.* 2022;146:110043. <https://doi.org/10.1016/j.ejrad.2021.110043>
11. Khoury ZH, Davis GE, Sultan AS. Artificial intelligence-enhanced virtual reality for pathology education. *Ann Diagn Pathol.* 2026;83:152639. <https://doi.org/10.1016/j.anndiagpath.2026.152639>
12. Hassell LA, Absar SF, Chauhan C, Dintzis S, Farver CF, Fathima S, et al. Pathology Education Powered by Virtual and Digital Transformation: Now and the Future. *Arch Pathol Lab Med.* 2023;147:474–91. <https://doi.org/10.5858/arpa.2021-0473-RA>

13. Prachyabrued M, Getrak M, Haddawy P, Sa-Ngasoongsong P, Zachmann G. Mixed Reality for Group-Based, Integrated Instruction of Anatomy, Pathology, and Physical Examination. 2026 IEEE Conf Virtual Real 3D User Interfaces Abstr Workshop VRW [Internet]. 2026 [cited 2026 June 28]. p. 931–5. <https://doi.org/10.1109/VRW70859.2026.00172>
14. Tsang HC, Morse RJ, Zhang J, Bui T, Emrich B, Hasan RA, et al. The clinical pathology laboratory in 360° virtual reality. *Am J Clin Pathol*. 2025;163:936–46. <https://doi.org/10.1093/ajcp/aqaf031>
15. Alcaraz-Mateos E, Abellan-Leante I, Galdame FS, Garcia-Hidalgo C, Perez-La-Encina M, Caballero-Aleman F, et al. Holopath-VR360-degree virtual reality videos in pathology education: The Holopath-VR experience. *Rev Esp Patol*. 2026;59:100861. <https://doi.org/10.1016/j.patol.2026.100861>
16. Khlaif Z, Salama N, Hamamra B, Mousa A. Factors Influencing Educators' Perspectives on Accepting Extended Reality in Health Care Education: Qualitative Study. *JMIR Med Educ*. JMIR Publications Inc., Toronto, Canada; 2025;11:e65042. <https://doi.org/10.2196/65042>
17. Taylor L, Dyer T, Al-Azzawi M, Smith C, Nzeako O, Shah Z. Extended reality anatomy undergraduate teaching: A literature review on an alternative method of learning. *Ann Anat - Anat Anz*. 2022;239:151817. <https://doi.org/10.1016/j.aanat.2021.151817>
18. Nebeling M, Rajaram S, Wu L, Cheng Y, Herskovitz J. XRStudio: A Virtual Production and Live Streaming System for Immersive Instructional Experiences. *Proc 2021 CHI Conf Hum Factors Comput Syst* [Internet]. New York, NY, USA: Association for Computing Machinery; 2021 [cited 2026 June 3]. p. 1–12. <https://doi.org/10.1145/3411764.3445323>
19. The jamovi project. jamovi [Internet]. 2026. <https://www.jamovi.org>
20. Veer V, Phelps C, Moro C. Incorporating Mixed Reality for Knowledge Retention in Physiology, Anatomy, Pathology, and Pharmacology Interdisciplinary Education: A Randomized Controlled Trial. *Med Sci Educ*. 2022;32:1579–86. <https://doi.org/10.1007/s40670-022-01635-5>
21. Kockro RA, Amaxopoulou C, Killeen T, Wagner W, Reisch R, Schwandt E, et al. Stereoscopic neuroanatomy lectures using a three-dimensional virtual reality environment. *Ann Anat Anat Anz Off Organ Anat Ges*. 2015;201:91–8. <https://doi.org/10.1016/j.aanat.2015.05.006>

Figures

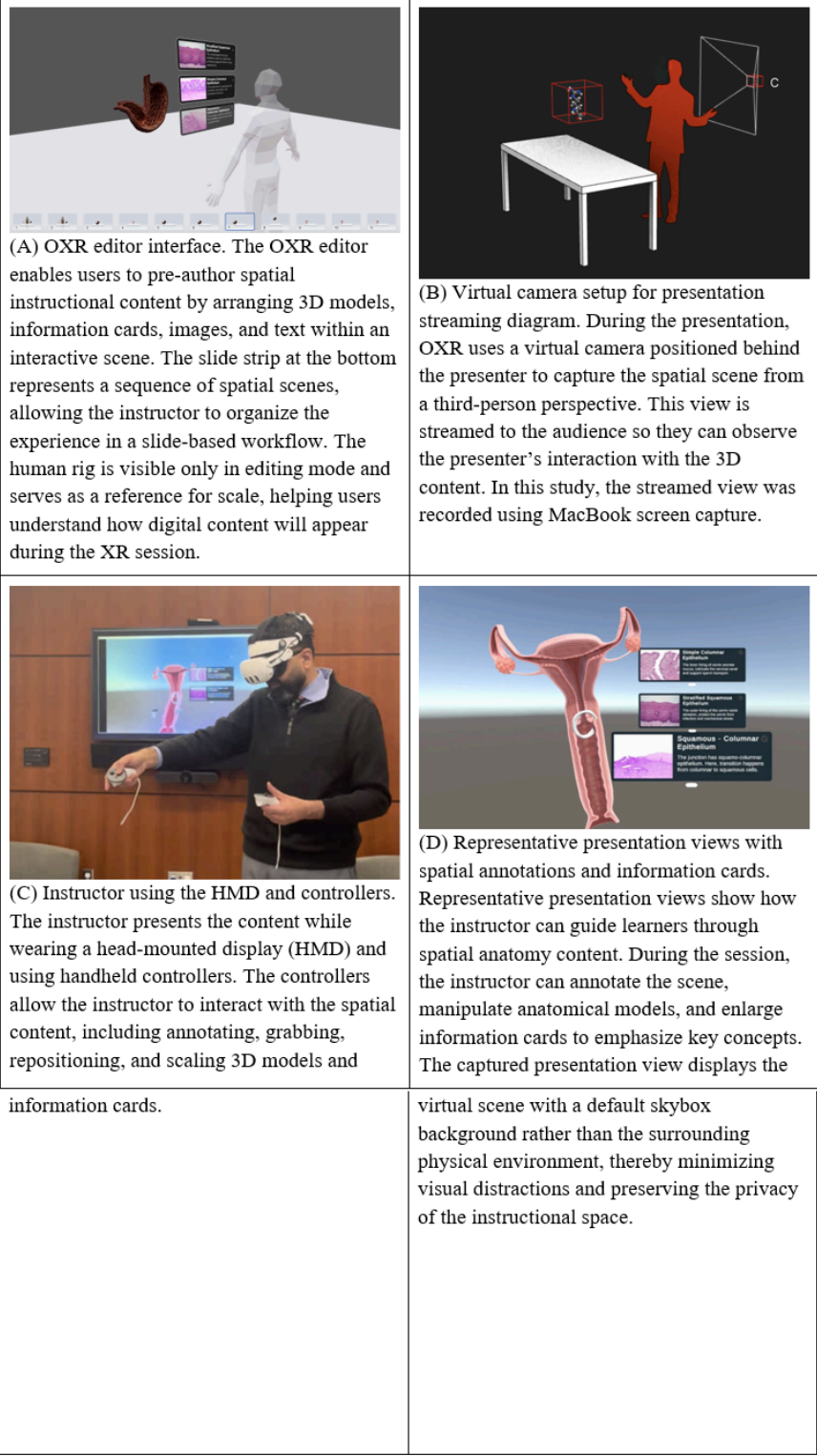


Figure 1

Overview of the OXR authoring and presentation workflow.

Supplementary Files

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

- [2dvideolecturecontent.pdf](#)
- [OXRvideolecturescriptcontent.pdf](#)