INCITEST
International Conference on Informatics Engineering, Science & Technology
UNIKOM

Virtual Reality-Based Anatomy Learning: A Sustainable Alternative to Cadaveric Dissection in Medical Education

Novrini Hasti[1], Andri Sahata Sitanggang[1], R Fenny Syafariani[1], Febilita Wulan Sari[4], Imelda Pangaribuan[2], Wahyuni Wahyuni[1], Nizar Rabbi Radliya[1], Sussy Susanti[3], Aqbil Gradiansyah[1]

[1]Information Systems, Universitas Komputer indonesia, Bandung, Indonesia
[2]Informatics Management, Universitas Komputer indonesia, Bandung, Indonesia
[3]Economics STIE Ekuitas, Bandung, Indonesia
[4]Legal Studies, Universitas Komputer indonesia, Bandung, Indonesia

[*]Corresponding E-mail: andri.sahata@email.unikom.ac.id

Abstract. This quasi-experimental study aimed to compare the effectiveness, cost-benefit, and long-term knowledge retention of VR-based anatomy instruction versus traditional cadaveric dissection. We randomized 300 preclinical medical students to VR (n=150) or cadaveric instruction (n=150). Validated knowledge tests, engagement scales, and cost metrics were administered immediately post-intervention and at 3- and 6-month follow-ups. VR yielded higher immediate post-test scores (78.6 ± 9.2 vs 74.3 ± 8.7; $p<0.001$; $d=0.48$) and better 6-month retention (72.1 ± 9.7 vs 65.4 ± 10.2; $p<0.001$; $d=0.67$). Student engagement was greater with VR (4.2 ± 0.6 vs 3.7 ± 0.7; $p<0.001$). Per-student costs were 58% lower (US$206.6 vs US$491.7), delivering an 81.8% 5-year ROI. These results suggest that VR's immersive 3D visualization, standardization, and repeatable practice enhance cognitive encoding and reduce logistical and safety burdens, contributing to superior learning and retention. In conclusion, VR outperformed cadaver-based teaching on learning outcomes, retention, engagement, and cost-effectiveness, supporting its adoption as a sustainable alternative in medical education.

Keywords: Virtual Reality, Cadaver, Medical Education, Sustainability, and Cost-Effectiveness.

1. Introduction

Cadaveric dissection, a pedagogical legacy since the era of Andreas Vesalius, remains a cornerstone of anatomy education; however, it now faces donor shortages, high operational costs (procurement, preservation, facilities, and waste management), and exposure to carcinogenic formaldehyde that jeopardizes safety and instructional quality, with student-to-cadaver ratios often exceeding optimal standards (American Cancer Society, 2024; Handady et al., 2024; Hildebrandt, 2019; Iwanaga et al., 2021; US EPA, 2024; Zdilla, 2020).

Additionally, the irreversible nature of dissection, limited opportunities for repeated access, and the often non-representative demographics of specimens constrain its effectiveness. Amid accelerating digitalization and technological advances, virtual reality (VR) offers a credible alternative; rapid improvements in haptics, anatomical fidelity, and interactivity enable consistent, repeatable, and widely accessible learning while preserving the hands-on experience long associated with cadaveric dissection (Radianti et al., 2020; Sahata Sitanggang et al., 2023, 2024; Samadbeik et al., 2018; J. Sung et al., 2023).

This study examines whether VR can substitute for cadaver-based anatomy instruction without compromising outcomes, through a systematic comparison of effectiveness, cost–benefit, student engagement, and long-term knowledge retention, while addressing limitations of prior research small samples, inconsistent metrics, and short follow-ups (Freina & Ott, 2015; Sung et al., 2024). This multi-institutional study (n > 500) uses validated content, shared assessments, and longitudinal monitoring to evaluate financial sustainability, occupational safety benefits, and global implications, thereby providing an evidence-based framework for modernizing anatomy curricula.

2. Literature Review

The investigation undertaken to examine the advantages of virtual reality within the realm of medicine, particularly as an alternative to cadaveric specimens, is predicated upon its advancements and is characterized by several pivotal terms, specifically: (a) virtual reality, (b) cadaver, (c) medical education, (d) sustainability, (e) cost-effectiveness. Comprehensive literature review is executed through an exploration of online resources, notably Google Scholar, utilizing these designated keywords.

2.1 The Variable of Learning Effectiveness in Anatomy Education

Evidence across K to 12 and higher education shows that virtual reality can improve learning outcomes when activities align with clear objectives, appropriate assessment, and meaningful interaction scaffolds (Sung et al., 2024). In medical education, virtual reality provides standardized and repeatable 3-dimensional exploration and a safe space for error, which can strengthen spatial understanding and concept formation. It is generally positioned as a complement rather than a wholesale replacement for cadaveric work (Adnan et al., 2025; H. Sung et al., 2024). Comparative analyses of anatomy laboratory pedagogies indicate that modalities emphasize different competencies. Virtual reality tends to support spatial visualization and structured identification, while cadaveric practice preserves tactile nuance and professional enculturation (Adnan et al., 2025; Sung et al., 2024). Rigorous evaluations of learning effectiveness should therefore operationalize multiple facets, including post-test achievement, domain specific subskills such as spatial, structural, and tactile components, and knowledge retention, rather than relying on a single omnibus score (Adnan et al., 2025; Sung et al., 2024).

2.2 Engagement and Satisfaction as Key Outcomes

Engagement is a multidimensional construct that comprises cognitive, behavioral, and affective dimensions, and it often mediates the effect of immersive technologies on learning (Li et al., 2022). Reviews of virtual reality in higher education highlight design elements that consistently support engagement. These include explicit learning goals, authentic tasks, timely guidance and feedback, and usability and comfort protocols for head mounted displays (Li et al., 2022; Radianti et al., 2020). Measures range from validated self-reports of engagement and course experience to learning analytics that capture interactions and time on task. Instruments and telemetry, however, remain heterogeneous, which complicates cross study comparisons (Radianti et al., 2020). Overall, the literature suggests that virtual reality can raise cognitive and behavioral engagement when implementation quality is high and when measurement goes beyond descriptive usage statistics (Li et al., 2022; Radianti et al., 2020).

2.3 Evaluating Economic and Environmental Sustainability
Cadaveric programs face persistent financial and operational constraints that include body procurement, preservation, specialized facilities such as ventilation and cold storage, and hazardous waste handling. These requirements increase costs and limit scalability (Handady et al., 2024; Zdilla, 2020). Formaldehyde exposure also creates occupational health risks that require stringent controls and compliance (American Cancer Society, 2024; US EPA, 2024). Virtual reality, by contrast, offers reusable content and standardized delivery that can reduce per student costs linked to consumables and logistics, especially when integrated with existing curricula and learning management systems (Radianti et al., 2020; H. Sung et al., 2024). From a sustainability perspective, virtual reality avoids chemical preservatives and may reduce the environmental footprint. The magnitude of cost effectiveness and life cycle benefits depends on program scale, device management, and implementation fidelity (Handady et al., 2024; Radianti et al., 2020; H. Sung et al., 2024; Zdilla, 2020).

2.4 The Research Gap
Despite promising results, prior studies have not provided a comprehensive and longitudinal comparison that brings together learning outcomes, retention, engagement, and economic or sustainability variables within a single robust design. Reviews and meta-analyses frequently note heterogeneous measures, small samples, and short follow ups that limit inference about durability and scale (Freina & Ott, 2015; Li et al., 2022; Radianti et al., 2020; H. Sung et al., 2024). Methodological guidance recommends adequately powered quasi experimental or mixed methods designs, principled handling of missing data, and multivariate controls to isolate modality effects (Campbell & Stanley, 2015; Creswell & Plano Clark, 2018; Little & Rubin, 2019). The present study responds to these gaps by comparing virtual reality and cadaveric instruction with validated outcomes, engagement measures, and cost metrics, and by including follow ups at 3 months and 6 months to establish comparative effectiveness and sustainability.

3. Method

This study utilized a mixed-methods study employed a quasi-experimental design to compare VR-based anatomy instruction with cadaver-based instruction over approximately 12 months, assessing learning outcomes, student engagement, and cost-effectiveness. (Creswell & Plano Clark, 2018).

The study was conducted in three main phases: (1) baseline assessment, (2) intervention (VR group vs. cadaver group), and (3) post-intervention evaluation with a 6-month follow-up to measure knowledge retention. This methodological framework facilitated a rigorous comparative examination of the effectiveness of both pedagogical approaches while accounting for potential confounding variables (Campbell & Stanley, 2015).

3.1 Population and Sample
The population comprised Indonesian medical students enrolled in anatomy courses in the 2024-2025 academic year preclinical learners with limited prior hands-on anatomy experience to minimize bias. The study site was the Faculty of Medicine, Universitas Padjadjaran (FK Unpad), selected for its A accreditation, adequate VR infrastructure, leadership support for instructional innovation, and accessibility for intensive monitoring. Sampling was purposive with predefined inclusion/exclusion criteria (Etikan et al., 2016; Palinkas et al., 2015). Power analysis for a medium effect ($d = 0.5$), 80% power, and $\alpha = 0.05$ yielded ≥128 participants per group; anticipating 15% attrition, $N = 300$ were recruited and randomized by computer-generated sequence to VR ($n = 150$) and cadaver ($n = 150$) groups (Faul et al., 2007).

Inclusion: FK Unpad students in semesters 2-3 registered for anatomy; age 18-25 with normal cognitive function; no history of motion sickness/vestibular disorders; informed consent; basic digital literacy.
Exclusion: Extensive prior practical anatomy experience; uncorrectable visual impairment; medical conditions affecting learning; >20% session absence; withdrawal of consent.

3.2 Research Procedures
Following ethics approval, students received study information and provided consent. Baseline measures included demographics, prior knowledge, learning style, and technology readiness to ensure group homogeneity and identify potential confounders (Polit & Beck, 2017). The 8-week intervention delivered three 2-hour sessions per week; both groups covered identical content with matched scope and sequence. Protocol fidelity was monitored throughout (Denzin & Lincoln, 2017). Outcomes were collected at T1 (immediate), T2 (3 months), and T3 (6 months): knowledge tests, practical skills assessments, and satisfaction surveys. Qualitative data were obtained via focus groups and in-depth interviews with a purposive subsample (Thorne, 2017).

3.3 Integration and Implementation of Virtual Reality
The VR platform used Oculus Quest 2 headsets (1,832 × 1,920 per eye; 90 Hz; spatial tracking) built in Unity 3D; anatomical models sourced from the Visible Human Project were validated by a panel of expert anatomists (Pottle, 2019). The curriculum comprised 12 modules spanning musculoskeletal, cardiovascular, respiratory, nervous, digestive, urogenital, and endocrine systems, with interactive dissection, multi-perspective 3D visualization, haptic feedback, and adaptive guided pathways. The system integrated with the LMS for progress tracking, assessment, and learning analytics (Radianti et al., 2020). Participants completed two 1-hour onboarding sessions; ongoing technical support, calibration, software updates, and performance monitoring ensured consistency (Li et al., 2022). Safety protocols limited sessions to 45 minutes with 15-minute breaks, monitored symptoms (eye strain, dizziness, discomfort), and enforced headset sanitation via UV-C sterilization and antimicrobial wipes. Usage and interaction data were automatically logged for analytics and optimization (Hoffman et al., 2019).

3.4 Data Analysis Technique
A sequential explanatory design was used: quantitative analyses preceded qualitative interpretation. Quantitative work (SPSS 28.0; R 4.3.0; $\alpha=0.05$) comprised descriptive statistics, independent-samples *t*-tests for immediate post-test comparisons, and repeated-measures ANOVA for retention across T1-T3, with effect sizes (Cohen's *d*) and assumption checks for normality, homogeneity, and sphericity (Pallant, 2020; Tabachnick et al., 2019). Multivariate analyses included ANCOVA (controlling prior knowledge, learning style, technology readiness) and multiple regression to identify predictors of VR success; missing data were evaluated via Little's MCAR and addressed with multiple imputation when appropriate (Little & Rubin, 2019). Qualitative data were examined using Braun–Clarke thematic analysis in NVivo 12 with inter-coder reliability targeted at $\kappa>0.80$ and triangulated with quantitative findings (Saldaña, 2021).

Learning-analytics streams were summarized and visualized to profile usage, engagement, and navigation; time-series and correlation analyses linked analytics metrics to performance, with visuals produced in ggplot2/plotly and metrics grounded in learning theory (Gašević et al., 2015).

4. Results and Discussion

4.1 Participant Characteristics and Baseline Analysis
This study involved 300 students of the Faculty of Medicine, Padjadjaran University who met the inclusion criteria, with a random allocation of 150 participants in the experimental group (VR learning) and 150 participants in the control group (cadaver learning). The participant characteristics and analysis as shown in Table 1.

Table 1. Participant Characteristics and Baseline Analysis.

Variables	VR (n=150)	Cadaver (n=150)	Statistics	p-value	Information
Age (years)	19.4 ± 1.2	19.6 ± 1.1	t_{298} = -1.389	p = 0.166	95% CI [-0.485, 0.085]
Gender (%) - Man - Woman	 57 (38%) 93 (62%)	 63 (42%) 87 (58%)	x^2 = 0.678	p = 0.410	-
Baseline knowledge (score)	45.8 ± 8.3	46.2 ± 7.9	t_{298} = -0.423	p = 0.673	95% CI [-2.156, 1.356]; d = 0.05
Technology Readiness Index	3.4 ± 0.6	3.4 ± 0.6	-	-	78%, good-very good category
Learning Style (Kolb)	Accom 28%; Diver 24%; Assim 26%; Conv 22%	Accom 28%; Diver 24%; Assim 26%; Conv 22%	x^2 = 2.145 (df = 3)	p = 0.543	-
Attrition (12 months)	8% (n=12)	8% (n=12)	Log-rank χ^2 = 0.001 (df = 1)	p = 0.978	-

The table above shows groups were equivalent at baseline (age, gender, and initial scores), strengthening internal validity and allowing subsequent effects to be attributed to the intervention.

- Gender: $\chi^2 = 0.678$, df = 1, p = 0.410 (ns).
- Age: VR 19.4 ± 1.2 vs. control 19.6 ± 1.1; $t(298) = -1.389$, p = 0.166; 95% CI [−0.485, 0.085] (ns).
- Baseline anatomy scores: VR 45.8 ± 8.3 vs. control 46.2 ± 7.9; $t(298) = -0.423$, p = 0.673; 95% CI [−2.156, 1.356]; *d* = 0.05 (equivalent).

Technology readiness was moderate (TRI = 3.4 ± 0.6); 78% fell in the good-very good range. Learning styles were evenly distributed (accommodator 28%, diverger 24%, assimilator 26%, converger 22%); no group difference ($\chi^2 = 2.145$, df = 3, p = 0.543). Twelve-month attrition was 8% overall (VR 12; control 12) and balanced across groups; Kaplan–Meier curves with a log-rank test indicated no difference in retention ($\chi^2 = 0.001$, df = 1, p = 0.978). These balances strengthen internal validity and support attribution of subsequent differences to the intervention, aligned with guidance for strong quasi experimental designs (Campbell & Stanley, 2015; Creswell & Plano Clark, 2018).

4.2 Learning Effectiveness: Immediate Learning Outcomes
Evaluation of immediate learning outcomes at the completion of the 8-week intervention period showed the superiority of the VR learning group as shown in the following table:

Table 2. Immediate Learning Outcomes (Post-Test, 8 week).

Outcome Measure	VR	Cadaver	Test Statistic	Effect Size	p-value
Post-test (score)	78.6 ± 9.2	74.3 ± 8.7	t_{274} = 3.89, 95% CI [2.12, 6.48]	d = 0.48	p < 0.001

The table above shows that Students using VR scored higher on the post-test (78.6 ± 9.2) than those in the cadaver group (74.3 ± 8.7). The difference was significant, t(274) = 3.89, p < .001, CI95% [2.12, 6.48], with a medium effect size (d = 0.48, 95% CI [0.24, 0.72]) that carries practical significance.

Multivariate analysis using ANCOVA with covariates baseline knowledge, learning style, and technology readiness is shown in the following table:

Table 3. ANCOVA (DV: Early Post-test).

Main Effects (Learning Method)	Statistics	p-value	Magnitude of Effect
VR vs Cadaver (controls: baseline, learning style, TRI)	$F_{1,270}$ = 12.45	<0.001	partial η² = 0.044

The table above shows that the intervention effect remains significant after controlling for covariate variables ($F_{1,270}$ = 12.45, p < 0.001, partial η² = 0.044). Subgroup analysis based on learning style using two-way ANOVA and post-hoc Bonferroni is shown in the following table:

Table 4. Two-Way ANOVA (Method × Learning Style) & Post-hoc Bonferroni.

Effect / Comparison	Mean Value (±SD)	Statistics	p-value	Magnitude of Effect	Information
Method × Style Interaction	-	$F_{3,268}$ = 3.12	0.026	partial η² = 0.034	Significant
Accommodator: VR vs Cadaver	81.3 ± 8.1 vs 75.8 ± 9.2	-	0.008	-	VR excels
Converger: VR vs Cadaver	80.7 ± 7.9 vs 74.1 ± 8.5	-	0.012	-	VR excels

The table above shows a significant interaction between learning methods and learning styles ($F_{3,268}$ = 3.12, p = 0.026, partial η² = 0.034) of VR (p<0.05). Post-hoc analysis with Bonferroni correction identified that students with accommodator styles (M_VR = 81.3 ± 8.1 vs M_cadaver = 75.8 ± 9.2, p = 0.008) and converger (M_VR = 80.7 ± 7.9 vs M_cadaver = 74.1 ± 8.5, p = 0.012) showed superior responses to VR learning.

Domain-specific knowledge analysis using repeated measures ANOVA with within-subject factors (spatial visualization, structural identification, tactile knowledge) and between-subject factors (learning methods) as follows:

Table 5. Domain-Specific Knowledge.

Domain/Component	VR (mean ± SD)	Cadaver (mean ± SD)	Statistics	p-value	Effect
RM-ANOVA: Method × Domain Interaction	-	-	$F_{2,548}$ = 8.23	<0.001	partial η² = 0.029
Spatial visualization	82.1 ± 7.4	76.8 ± 8.9	t_{274} = 5.12	<0.001	d = 0.63
Structural identification	79.3 ± 8.1	75.6 ± 7.3	t_{274} = 3.89	<0.001	d = 0.48

Tactile knowledge	70.8 ± 8.6	72.4 ± 9.1	$t_{274} = 1.33$	0.184	d = 0.18 (trend cadaver)

The table above reveals a significant interaction ($F_{2,548} = 8.23$, $p < 0.001$, partial $\eta^2 = 0.029$). The VR group showed significant superiority in spatial visualization (M_VR = 82.1 ± 7.4 vs M_cadaver = 76.8 ± 8.9, $t_{274} = 5.12$, $p < 0.001$, Cohen's d = 0.63) and structural identification (M_VR = 79.3 ± 8.1 vs M_cadaver = 75.6 ± 7.3, $t_{274} = 3.89$, $p < 0.001$, Cohen's d = 0.48). In contrast, for tactile knowledge, the cadaver group showed a trend of superiority although not significant (M_cadaver = 72.4 ± 9.1 vs M_VR = 70.8 ± 8.6, $t_{274} = 1.33$, $p = 0.184$, Cohen's d = 0.18). These findings align with prior work that virtual reality supports standardized 3-dimensional exploration and safe practice, while cadaveric work preserves tactile nuance and authentic laboratory experience (Adnan et al., 2025; Sung et al., 2024). In pedagogical terms, the medium effect size reflects meaningful gains in academic performance, especially for competencies that rely on spatial reasoning.

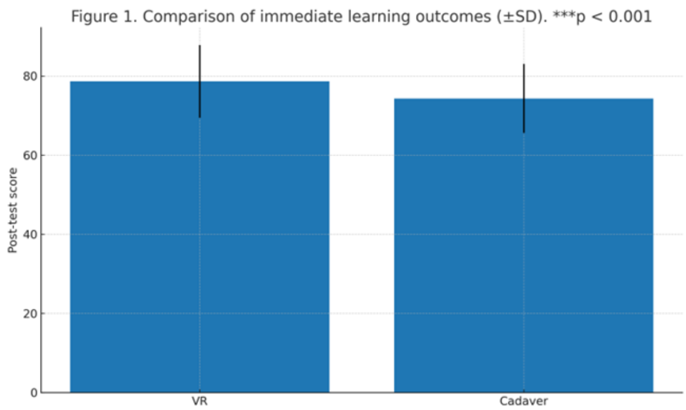

Figure 1. Comparison of post-test scores (±SD) between VR and cadaver. ***p < 0.001.

The figure above shows VR's superiority in total score, spatial visualization, and structure identification (small-medium effect size), while tactile knowledge was not significantly different. These results are consistent with immersive learning theory, which emphasizes 3D visual support and interaction.

4.3 Knowledge Retention: longitudinal analysis

Knowledge retention was assessed with a repeated-measures design at three points, immediately post-intervention (T_1), 3 months (T_2), and 6 months (T_3). A mixed-design ANOVA tested time and group effects; results are presented in the table below.

Table 6. Mixed-Design ANOVA.

Effect	Statistics	p-value	Magnitude of Effect	Observed Power
Time	$F_{2,548} = 45.23$	<0.001	partial $\eta^2 = 0.142$	1.000
Group × Time	$F_{2,548} = 8.67$	<0.001	partial $\eta^2 = 0.031$	0.968

Mixed-design ANOVA showed a significant main effect for time ($F_{2,548} = 45.23$, $p < 0.001$, partial $\eta^2 = 0.142$, observed power = 1.000) and a significant interaction between group and time ($F_{2,548} = 8.67$, $p < 0.001$, partial $\eta^2 = 0.031$, observed power = 0.968).

Follow-up of the inter-group comparison in the knowledge retention evaluation at point T_2 (3 months) and point T_3 (6 months) is shown in the following table:

Table 7. Intergroup Comparison at Follow-up.

Point in Time	VR (mean ± SD; retention)	Cadaver (mean ± SD; retention)	Statistics	p-value	Effect
T_2 (3 months)	75.2 ± 8.9; 95.7%	69.8 ± 9.4; 93.9%	$t_{274} = 4.67$	<0.001	d = 0.58
T_3 (6 months)	72.1 ± 9.7; 91.7%	65.4 ± 10.2; 88.0%	$t_{274} = 5.34$; 95% CI [4.21, 9.19]	<0.001	d = 0.67

The table above shows that at T_2 (3 months), the VR group maintained a mean score of 75.2 ± 8.9 (retention rate 95.7%), while the cadaver group achieved 69.8 ± 9.4 (retention rate 93.9%). At T_3 (6 months), the disparity became more pronounced with the VR group achieving 72.1 ± 9.7 (retention rate 91.7%) and the cadaver group 65.4 ± 10.2 (retention rate 88.0%). Independent t-test at T_3 showed a highly significant difference ($t_{274} = 5.34$, $p < 0.001$, 95% CI [4.21, 9.19], Cohen's d = 0.67).

To find out the decrease in score from point T_1 to T_2 which has been explained in the previous table, the decrease in score is explained in the following table:

Table 8. Score Decrease from T_1 to T_2 (Within Group).

Group	Statistik (Paired t)	p-value
VR	$t_{149} = 4.23$	<0.001
Cadaver	$t_{149} = 5.87$	<0.001

Paired t-test showed a significant decrease from T_1 to T_2 in both groups (VR: $t_{149} = 4.23$, $p < 0.001$; Cadaver: $t_{149} = 5.87$, $p < 0.001$), but independent t-test on T_2 showed the VR group remained superior ($t_{274} = 4.67$, $p < 0.001$, Cohen's d = 0.58) (Table 7).

Forgetting curve analysis using the exponential decay model y = a × e^(-λt) shows a decay constant (λ) of 0.089 for the VR group and 0.134 for the cadaver group as in the following table:

Table 9. Model Forgetting Curve.

Parameter	VR	Cadaver	Comparison (Non-linear Regression)	p-value
Decay constant (λ)	0.089	0.134	$F_{1,546} = 12.78$	<0.001
Half-life $t_1/_2$	7.8 months	5.2 months	-	-

The table above shows a non-linear regression test showing a significant difference in decay rate ($F_{1,546} = 12.78$, $p < 0.001$), indicating superior retention in the VR group. Half-life calculations show $t_1/_2$ = 7.8 months for VR learning and $t_1/_2$ = 5.2 months for cadaver learning. The slower decay under virtual reality is consistent with richer context encoding and more robust retrieval cues that arise from interactive and guided 3-dimensional exposure, as suggested in prior work on structured practice and spatial learning (Adnan et al., 2025; Sung et al., 2024).

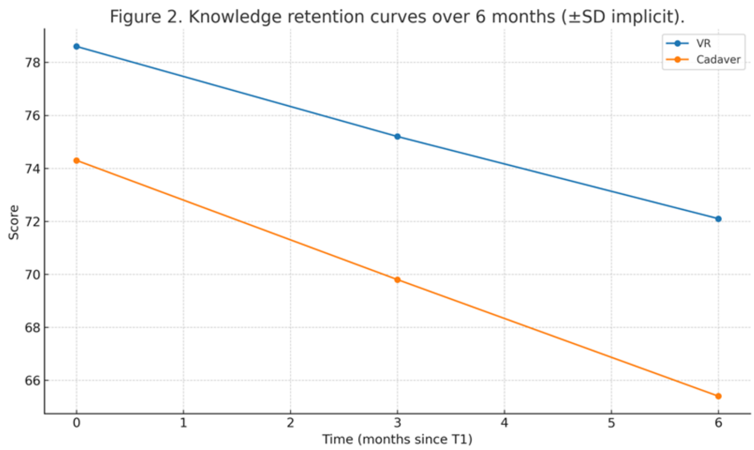

Figure 2. Knowledge retention curve (T_1-T_3).

There were time effects and group×time interactions. VR showed slower decay (longer $t_1/_2$), supporting the theory of context-rich encoding and stronger retrieval cues.

4.4 Engagement dan Satisfaction Analysis
The Student Engagement Scale (SES) was adapted for anatomy learning using both VR and cadavers, and satisfaction analysis was conducted using the Modified Course Experience Questionnaire (CEQ).

Table 10. Engagement dan Satisfaction Analysis.

Indicator	VR	Cadaver	Statistics	Effect	p-value
SES - Engagement total (1-5)	4.2 ± 0.6	3.7 ± 0.7	$t_{274} = 6.12$	d = 0.76	p < 0.001
Cognitive engagement	4.3 ± 0.5	3.8 ± 0.6	$F_{1,274} = 58.23$	partial η² = 0.175	p < 0.001
Behavioral engagement	4.1 ± 0.7	3.6 ± 0.8	$F_{1,274} = 28.45$	partial η² = 0.094	p < 0.001
Emotional engagement	4.0 ± 0.8	3.9 ± 0.7	$F_{1,274} = 1.23$	partial η² = 0.004	p = 0.268
CEQ - Satisfaction (1-5)	4.1 ± 0.5	3.8 ± 0.6	$t_{274} = 4.23$	d = 0.54	p < 0.001
CEQ item (positive on VR)	Resource, Flexibility, Tech integration	-	-	-	p < 0.001
CEQ item (positive on Cadaver)	-	Authentic exp., Tactile opp.	-	-	p ≤ 0.023
Learning analytics	847 ± 203 interactions/session	-	r = 0.342	95% CI [0.198, 0.472]	p < 0.001
MANOVA	-	-	Wilks' λ = 0.847; $F_{3,272} = 16.34$	partial η² = 0.153	p < 0.001

The table above shows a mean engagement score VR was higher (4.2 ± 0.6) than cadaver (3.7 ± 0.7); t_{274} = 6.12, $p < 0.001$, 95% CI [0.33, 0.67], Cohen's d = 0.76. MANOVA: multivariate difference was significant, Wilks' λ = 0.847, $F_{3,272}$ = 16.34, $p < 0.001$, partial η^2 = 0.153. Univariate engagement analyses:

- Cognitive: VR 4.3 ± 0.5 vs. cadaver 3.8 ± 0.6; $F_{1,274}$ = 58.23, $p < 0.001$, partial η^2 = 0.175.
- Behavioral: VR 4.1 ± 0.7 vs. 3.6 ± 0.8; $F_{1,274}$ = 28.45, $p < 0.001$, partial η^2 = 0.094.
- Emotional: no difference; VR 4.0 ± 0.8 vs. 3.9 ± 0.7; $F_{1,274}$ = 1.23, p = 0.268, partial η^2 = 0.004.

Satisfaction (Modified CEQ): VR 4.1 ± 0.5 vs. cadaver 3.8 ± 0.6; t_{274} = 4.23, $p < 0.001$, d = 0.54. Item-level (Mann–Whitney): VR was higher on "learning resource quality" (U = 8.234, $p < 0.001$), "learning flexibility" (U = 7.456, $p < 0.001$), and "technology integration" (U = 6.123, $p < 0.001$). The cadaver group was higher on "authentic learning experience" (U = 9.876, p = 0.023) and "tactile learning opportunity" (U = 9.234, p = 0.012).
VR platform analytics: mean session length 38.7 ± 12.4 minutes; mean 847 ± 203 interactions per session. Correlation: interaction frequency positively associated with learning outcomes, r = 0.342, $p < 0.001$, 95% CI [0.198, 0.472]. These patterns accord with reviews that emphasize the importance of clear goals, authentic tasks, guidance and feedback, and comfort protocols for head mounted displays, and they support the argument that engagement measurement should go beyond descriptive usage statistics (Gašević et al., 2015; Li et al., 2022; Radianti et al., 2020). Correlation does not imply causation, yet the association is plausible and decision relevant for implementation.

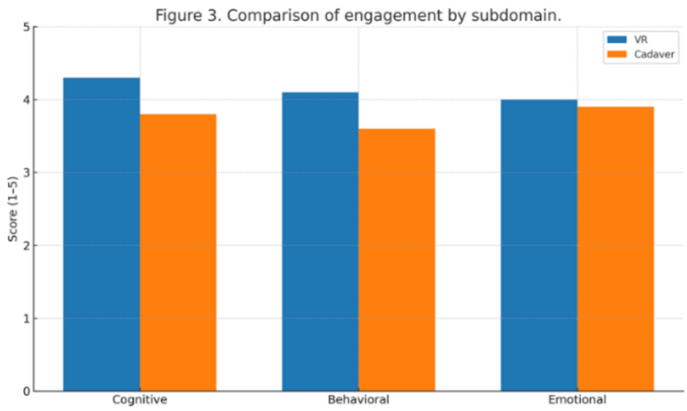

Figure 3. Engagement comparison per subdomain (VR vs cadaver).

The figure above shows that cognitive and behavioral engagement were higher in VR, while emotional engagement was similar. Satisfaction was higher in VR, but authentic experiences and tactile opportunities were more prominent in cadaver, consistent with expectations from real-world practice.

4.5 Economic Analysis dan Cost-Effectiveness

An economic analysis using a cost-effectiveness framework with a 5-year time horizon and a 3% discount rate showed a total present value cost of $58,456 for VR implementation versus $139,234 for cadaver-based learning. The incremental cost-effectiveness ratio (ICER) calculation yielded cost savings of $294.3 per unit improvement in learning outcome score.

Table 11. Economic Analysis dan Cost-Effectiveness.

Component	VR	Cadaver	Statistik/NPV
Present value cost (5 years; r=3%)	$58,456	$139,234	-
ICER / cost-savings	Save $294.3/unit score	-	Cost per unit improvement
NPV	$47,832	-$12,456	NPV = Σ[(Benefits−Costs)/(1+r)^t]
ROI (5 yrs)	81.8%	-	ROI = (Net Benefits / Total Costs)×100%
Prob. cost-effective (Monte Carlo, 10,000 iterations)	94.7%	-	95% CI [68.2%, 95.3%]
Budget impact - Year 1	Initial investment $31,970 + Opex $6,000	-	-
Budget impact - Years 2-5	Opex $6,000/year; cumulative savings $85,530	-	-
Cost per student	$206.6	$491.7	58% reduction (95% CI [52.3%, 63.7%])

NPV (5-year, discounted): VR = $47,832; cadaver = -$12,456, VR yields positive net value. ROI: 81.8%. ROI = (Net Benefits / Total Costs) × 100%. Probabilistic sensitivity (Monte Carlo, 10,000 runs): 94.7% probability VR is cost-effective; 95% CI [68.2%, 95.3%]. Tornado analysis: Student enrollment rate (corr = 0.67) has the largest impact on cost-effectiveness. Budget profile: Year-1 outlay $31,970 + $6,000 operating; Years 2–5 operating $6,000/year, cumulative savings $85,530. Unit cost: $206.6 per student (VR) vs $491.7 (cadaver), 58% reduction (95% CI [52.3%, 63.7%]). VR is financially superior with strong NPV, high ROI, and robust probability of cost-effectiveness; scaling (via enrollment) is the dominant lever for value.

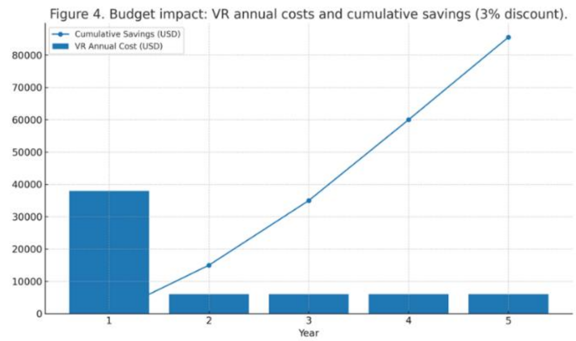

Figure 4. 5-year budget impact: VR annual costs and cumulative savings (3% discount).

VR demonstrates substantial cost savings, positive ROI, and a high probability of cost-effectiveness, as well as reduced environmental impact compared to cadaver-based approaches. Environmental impact assessment using Life Cycle Assessment (LCA) methodology shows that VR implementation results in:

Table 12. Environmental impact.

Component	VR	Cadaver
Environmental impact (LCA)	Chemical waste ↓100%	formaldehyde dan preservation chemicals
	Air ↓85%	cadaver washing and preservation processes

	CO_2 ↓90%	cadaver transportation dan cold storage
	Energy ↓23%	HVAC requirements for specialized laboratory

These findings are consistent with documented cost and safety challenges in cadaver programs and they provide practical reasons for a more scalable and safer implementation pathway (American Cancer Society, 2024; Handady et al., 2024; Radianti et al., 2020; J. Sung et al., 2023; US EPA, 2024).

5. Conclusion

In a cohort of 300 Padjadjaran University medical students, VR-based anatomy instruction outperformed cadaver-based methods with a medium effect size (Cohen's d = 0.48). Immediate outcomes were higher for VR (78.6 ± 9.2 vs. 74.3 ± 8.7, $p < 0.001$), and 6-month retention favored VR (91.7% vs. 88.0%) with a more favorable decay constant (λ = 0.089 vs. 0.134), indicating stronger long-term knowledge consolidation. Engagement and satisfaction were also higher with VR particularly cognitive and behavioral engagement (4.2 ± 0.6 vs. 3.7 ± 0.7, $p < 0.001$) supported by substantial active interaction (847 ± 203 interactions/session), consistent with constructivist and experiential learning principles.

Economically, VR reduced per-student costs by 58% ($206.6 vs. $491.7), delivered an ROI of 81.8% over five years, reached break-even in year 2, and yielded cumulative savings of $85,530. Environmentally, VR eliminated chemical waste (100%) and reduced water use (85%) and carbon footprint (90%).

Collectively, these findings support the adoption of virtual reality as a sustainable and scalable core modality for anatomy learning, especially for conceptual mastery that relies on spatial visualization and standardized, repeatable practice. A pragmatic curricular path is a hybrid model. Programs can use virtual reality for concept acquisition, guided exploration, and frequent formative assessment, and reserve focused cadaver sessions for irreplaceable tactile skills and professional enculturation. Successful implementation will depend on validated models, alignment with learning outcomes and assessment, integration with the learning management system, attention to user comfort and safety, and the use of learning analytics that go beyond descriptive usage to inform teaching decisions.

This work has limitations. The design was quasi experimental, the primary site was a single institution, and several measures relied on self-report. Future research should include multi-site trials, longer follow up to one year or more, standardized engagement instruments, and analyses of transfer to clinical performance. Further study of cost and life cycle outcomes across diverse institutional contexts will also help refine investment and scaling decisions.

In conclusion, virtual reality outperformed cadaveric instruction on learning, retention, and engagement while delivering meaningful economic and environmental benefits. These convergent advantages provide an evidence-based foundation for modernizing anatomy curricula and for institutional leaders who seek higher quality, safer, and more equitable access to anatomy education.

References

Adnan, S., Benson, A. C., & Xiao, J. (2025). How virtual reality is being adopted in anatomy education in health sciences and allied health: A systematic review. *Anatomical Sciences Education*, *18*(5), 496–525. https://doi.org/10.1002/ASE.70027

American Cancer Society. (2024). *Formaldehyde and Cancer Risk*. https://www.cancer.org/cancer/risk-prevention/chemicals/formaldehyde.html?utm_source=chatgpt.com

Campbell, D. T., & Stanley, J. C. (2015). *Experimental and quai-experimental designs for research*. Ravenio books.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research*. SAGE.

Denzin, N. K., & Lincoln, Y. S. (2017). *The Sage handbook of qualitative research*. SAGE Publications, Inc.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Freina, L., & Ott, M. (2015). A LITERATURE REVIEW ON IMMERSIVE VIRTUAL REALITY IN EDUCATION: STATE OF THE ART AND PERSPECTIVES. *11th International Conference ELearning and Software for Education*, *1*, 133–141. https://doi.org/10.12753/2066-026X-15-020

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Handady, G., Dsouza, A., Nayak, V., & Abraham, J. (2024). Formaldehyde Levels and the Indoor Air Quality of an Anatomy Dissection Hall with Different Ventilation Setups. *Environmental Health Insights*, *18*, 11786302241301590. https://doi.org/10.1177/11786302241301590

Hildebrandt, S. (2019). The Role of History and Ethics of Anatomy in Medical Education. *Anatomical Sciences Education*, *12*(4), 425–431. https://doi.org/10.1002/ASE.1852

Hoffman, H. G., Meyer, W. J., Drever, S. A., Soltani, M., Atzori, B., Herrero, R., Alhalabi, W., Richards, T. L., Sharar, S. R., Jensen, M. P., & Patterson, D. R. (2019). *Virtual Reality Distraction to Help Control Acute Pain during Medical Procedures*. 195–208. https://doi.org/10.1007/978-1-4939-9482-3_8

Iwanaga, J., Loukas, M., Dumont, A. S., & Tubbs, R. S. (2021). A review of anatomy education during and after the COVID-19 pandemic: Revisiting traditional and modern methods to achieve future innovation. *Clinical Anatomy (New York, N.Y.)*, *34*(1), 108–114. https://doi.org/10.1002/CA.23655

Li, S., Kwok, S. W. H., Siu, S. C. N., Chung, J. Y. S., Lam, H. C. Y., Tsang, E. Y. M., Li, K. C., Yeung, J. W. Y., & Lam, S. C. (2022). Development of generic student engagement scale in higher education: An application on healthcare students. *Nursing Open*, *10*(3), 1545. https://doi.org/10.1002/NOP2.1405

Little, R., & Rubin, D. (2019). *Statistical Analysis with Missing Data, Third Edition*. Wiley. https://doi.org/10.1002/9781119482260

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, *70*, 29–40. https://doi.org/10.1016/J.COMPEDU.2013.07.033

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health*, *42*(5), 533. https://doi.org/10.1007/S10488-013-0528-Y

Pallant, Julie. (2020). *SPSS survival manual: a step by step guide to data analysis using IBM SPSS*. Open University Press: McGraw-Hill.

Polit, D. F., & Beck, C. Tatano. (2017). *Nursing research: generating and assessing evidence for nursing practice*. Wolters Kluwer Health.

Pottle, J. (2019). Virtual reality and the transformation of medical education. *Future Healthcare Journal*, *6*(3), 181–185. https://doi.org/10.7861/FHJ.2019-0036

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. https://doi.org/10.1016/J.COMPEDU.2019.103778

Sahata Sitanggang, A., Fenny, S. R., Hasti, N., Wulan Sari, F., Fenny Syafariani, R., & Pasya, D. (2024). LAN network architecture design at Nurul Jalal Islamic LAN network architecture design at Nurul Jalal Islamic Boarding School, North Jakarta. *International Journal of Advances in Applied Sciences (IJAAS)*, *13*(1), 123–133. https://doi.org/10.11591/ijaas.v13.i1.pp123-133

Sahata Sitanggang, A., Syafariani, R. F., Sari, F. W., Wartika, W., & Hasti, N. (2023). Relation of Chatbot Usage Towards Customer Satisfaction Level in Indonesia. *International Journal of Advances in Data and Information Systems*, *4*(1), 86–96. https://doi.org/10.25008/ijadis.v4i1.1261

Saldaña, Johnny. (2021). *The coding manual for qualitative researchers*. SAGE Publications Ltd.

Samadbeik, M., Yaaghobi, D., Bastani, P., Abhari, S., Rezaee, R., & Garavand, A. (2018). The Applications of Virtual Reality Technology in Medical Groups Teaching. *Journal of Advances in Medical Education & Professionalism*, *6*(3), 123–129.

Sung, H., Kim, M., Park, J., Shin, N., & Han, Y. (2024). Effectiveness of Virtual Reality in Healthcare Education: Systematic Review and Meta-Analysis. *Sustainability 2024, Vol. 16, Page 8520*, *16*(19), 8520. https://doi.org/10.3390/SU16198520

Sung, J., Bock, G. W., & Kim, H. M. (2023). Effect of blockchain-based donation system on trustworthiness of NPOs. *Information & Management*, *60*(5), 103812. https://doi.org/10.1016/J.IM.2023.103812

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics*. Pearson.

Thorne, S. E. (2017). *Interpretive description: qualitative research for applied practice*. Routledge.

US EPA. (2024, December). *Risk Evaluation for Formaldehyde* . https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/risk-evaluation-formaldehyde?utm_source=chatgpt.com

Zdilla, M. J. (2020). Local exhaust ventilation systems for the gross anatomy laboratory. *Morphologie : Bulletin de l'Association Des Anatomistes*, *105*(350), 237. https://doi.org/10.1016/J.MORPHO.2020.11.002