Morphologie 110 (2026) 101125

Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

ORIGINAL ARTICLE

Instrumenting augmented reality in anatomy and histology education: What student usage reveals about engagement, motivation and learning in European and Sub-Saharan African contexts

D. Mutombo Mwembo[a,b], A. Fries[b], P. Bonnet[b], A. Sul Ilunga Sul[a], O. Prygiel[b], A. Lagier[b], W. Arung Kalau[a], S. Sobczak[c], M. Schyns[d], V. Defaweux[b,*]

[a] *Department of Surgery, Faculty of Medicine, University Clinics of Lubumbashi, University of Lubumbashi, Lubumbashi, Democratic Republic of the Congo*
[b] *Human Anatomy Laboratory, Department of Biomedical and Preclinical Sciences, Faculty of Medicine, University of Liège, Liège, Belgium*
[c] *Department of Anatomy, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada*
[d] *HEC Liège, Management School of the University of Liège, Liège, Belgium*

KEYWORDS
Augmented reality;
Anatomy education;
Histology education;
Learning analytics;
Visuospatial skills;
Student engagement

Summary The effectiveness of augmented reality (AR) in anatomy and histology education remains variable across literature, suggesting strong dependence on pedagogical context, learner characteristics, and conditions of use. ARanatomy was developed as a mobile AR-learning environment to support spatial understanding of anatomical structures across multiple levels of organization, from macroscopic to histological, and required contextualized evaluation under authentic curricular conditions. This multicentric study involved 2,031 students across five cohorts from three institutions across Europe and Sub-Saharan Africa. Three successive versions of the application were deployed, differing in ergonomic maturity and level of usage instrumentation. Collected data included student perceptions (WBLT framework), motivation (Self-Determination Theory), objective usage traces, and performance outcomes when available. Descriptive, comparative, and correlational analyses were conducted, and usage profiles were identified using *K*-means clustering ($k = 3$). Perceptions of learning and design varied markedly across cohorts. Engagement differences between users and non-users were limited. Instrumented analyses showed that intensive and diversified use was associated with higher

* Corresponding author.
E-mail address: Valerie.defaweux@uliege.be (V. Defaweux).

https://doi.org/10.1016/j.morpho.2026.101125

visuospatial performance, whereas no effect was observed on immediate disciplinary performance. Usage intensity correlated positively with intrinsic motivation and positive emotions and negatively with amotivation. These findings indicate that the educational value of ARanatomy depends primarily on its pedagogical integration, ergonomic maturity, and depth of use. AR was associated with higher visuospatial performance than with short-term academic performance, although no causal relationship could be established. Taking together, these results highlight the importance of analytics-informed and pedagogically aligned implementation in anatomy education.

Introduction

Over the past decade, extended reality technologies — including virtual reality (VR), augmented reality (AR), and hybrid environments — have increasingly been integrated into biomedical education. Systematic reviews highlight their potential to enhance student engagement, improve understanding of three-dimensional anatomical structures, and enrich learning experiences [1,2]. These tools appear particularly relevant in anatomy and histology, where learners must integrate spatial relationships across multiple levels of organization, from the macroscopic to the ultrastructural.

Despite this growing interest, empirical findings remain heterogeneous. While some studies report learning gains, others find little or no effect on immediate academic performance. This variability affects both knowledge-based outcomes and visuospatial abilities, which are often considered primary targets of immersive technologies [3,4]. Such discrepancies raise a central question: do these technologies enhance learning *per se*, or are their effects largely contingent upon the context in which they are implemented?

Accumulating evidence suggests that effectiveness is strongly conditioned by pedagogical integration. Factors such as instructional design, learner guidance, technological maturity, ergonomic stability, and cognitive load substantially modulate observed outcomes [5,6]. Consequently, AR and VR-based learning environments should not be viewed as uniform interventions but rather as tools whose impact depends on how they are embedded within educational contexts.

Beyond pedagogical design, individual learner characteristics play a critical role. Visuospatial abilities, prior digital experience, and motivational profiles influence adoption and depth of use [3,7,8]. Students with higher intrinsic motivation or stronger spatial skills tend to benefit more from three-dimensional representations, whereas amotivation and limited digital fluency are associated with superficial engagement and reduced learning benefits. These factors contribute to pronounced intra- and inter-cohort variability.

This dual dependence — on implementation context and learner characteristics — poses important methodological challenges. Few studies articulate macro-level curricular conditions, meso-level engagement patterns, and micro-level cognitive or motivational mechanisms within a single analytical framework. Moreover, most evaluations rely primarily on perception surveys, which provide limited insight into actual usage behaviors. Learning-analytics research consistently shows discrepancies between declared and observed use, underscoring the need for objective instrumentation [9,10].

Another limitation of literature lies in the predominance of mono-centric studies conducted within highly homogeneous academic settings. Such approaches restrict generalizability and complicate the distinction between tool effects and contextual influences. Multicentric designs, by contrast, allow exploration of how institutional, cultural, and pedagogical variability modulates both usage patterns and learning outcomes.

Finally, studies combining perception, motivation, instrumented usage traces, and performance outcomes remain rare. Yet this integration is essential for understanding how AR- and VR-based learning environments are appropriate in authentic settings: which forms of use correspond to which learner profiles, how motivation shapes engagement, and under what conditions immersive technologies support visuospatial skills or disciplinary learning.

Responding to these gaps, the present multicentric study examines five cohorts from three institutions across diverse pedagogical contexts in anatomy and histology. It deploys successive versions of the same AR environment, differing in their level of instrumentation, enabling an integrated analysis of perceptions, motivation, engagement, usage behaviors, and performance. This design provides a robust framework for investigating how contextual conditions and individual differences shape the educational impact of AR.

Based on existing literature and preliminary observations, we formulated three hypotheses: (H1) the use of AR is positively associated with student engagement; (H2) usage intensity and modality are associated with differentiated motivational and engagement profiles; and (H3) the effects of AR-based learning environments on performance depend on the pedagogical context and the type of competence assessed. The specific application used in this study is described in detail in the Materials and Methods section.

Materials and methods

ARanatomy: application design and technopedagogical framework

ARanatomy is a mobile AR application designed to support learning in anatomy and histology through interactive three-dimensional representations displayed on a screen-

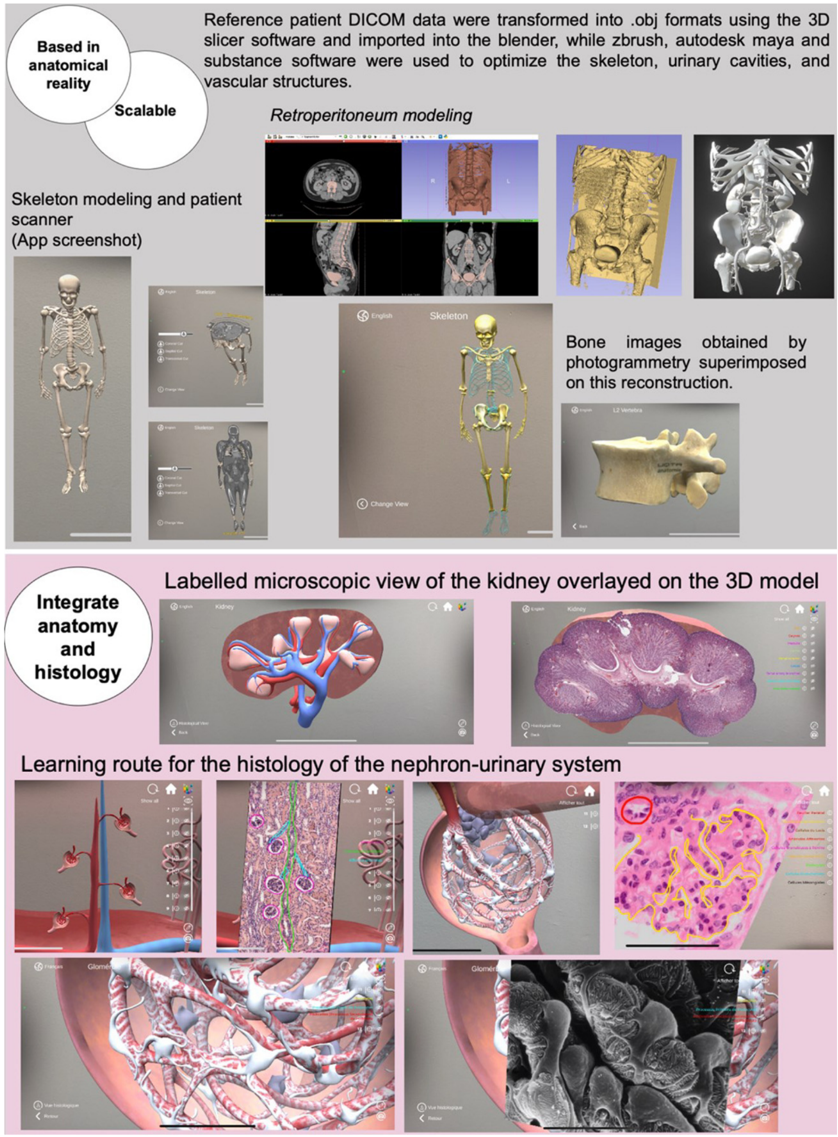

Figure 1 Multiscale anatomical modeling and integration of histological levels in ARanatomy. The first panel illustrates the generation of anatomically accurate 3D models from patient-derived DICOM data, including segmentation and reconstruction of the skeletal framework, urinary cavities, and vascular structures, complemented by photogrammetry-based bone data. The second panel shows the integration of histological information, with superimposition of optical and electron microscopy images onto the anatomical models, enabling continuity between macroscopic organization and ultrastructural components. The application is freely available for download at the following links: https://play.google.com/store/apps/details?id=com.ULiege.ARAnatomy&hl (Android version) and https://apps.apple.com/be/app/arnatomy-hec/id1523284550 (iOS version).

based interface, without full sensory immersion as defined in virtual reality environments. The application integrates anatomically accurate 3D models reconstructed from medical imaging (DICOM) data, complemented by histological content enabling navigation across multiple levels of biological organization, from macroscopic structures to ultrastructural components. In specific modules, optical and electron microscopy images are overlaid onto digital reconstructions to enhance spatial coherence (Fig. 1).

From its inception, ARanatomy was conceived as an adaptable pedagogical environment suitable for both supervised and autonomous learning contexts. Its development followed an iterative process involving educators, students, and developers, leading to successive refinements targeting interface stability, interaction fluidity, and cognitive ergonomics. Student feedback from early deployments informed major redesigns addressing manipulation difficulties, navigation clarity, and integration of learning pathways.

Instrumentation of usage and version evolution

Three levels of usage instrumentation were implemented across successive versions of the application:

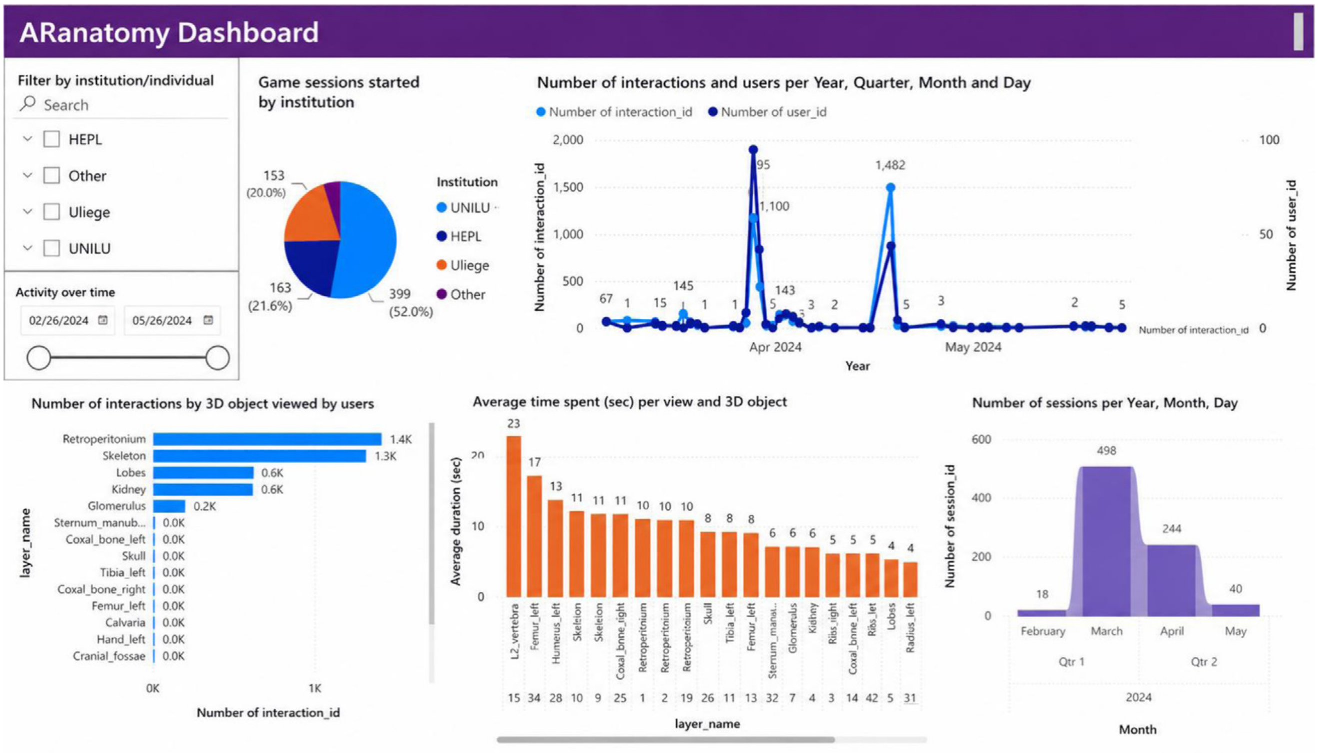

Figure 2 Dashboard of user interactions within the ARanatomy application. Interactive dashboard summarizing user interactions within the ARanatomy application. The dashboard displays temporal patterns of use, distribution of sessions and interactions across institutions, and indicators of content exploration (3D objects viewed, session frequency, and average interaction duration).

- version 0: non-instrumented use, without automated data collection;
- version 1: partial instrumentation, capturing session counts and accessed views;
- version 2: advanced instrumentation, including detailed metrics on AR interactions, exploration of histological layers, interaction duration, and navigation patterns across modes.

In instrumented versions of the application, usage data were automatically collected and visualized using Microsoft Power BI [11].

Based on these logs, synthetic indicators were computed to characterize usage intensity and depth (Fig. 2). To identify natural usage profiles, an unsupervised k-means clustering approach ($k = 3$) was applied, with the choice of k guided by considerations of interpretability and cluster stability, yielding three categories: low users, moderate users, and high users. These profiles were then used to examine relationships between usage behavior, motivation, perception, and performance.

In parallel, the application underwent a progressive process of ergonomic refinement across versions. While version 0 was characterized by notable usability limitations, subsequent iterations led to increasingly stable, structured, and user-friendly interactions. version 1 introduced initial improvements in interaction consistency and navigation, whereas version 2 reflects a more mature ergonomic design, with smoother navigation, standardized gestures, enhanced visual transitions, and improved integration of layered anatomical and histological content.

Study context and description of the cohorts

The study included 2,031 students distributed across five cohorts from three higher education institutions, encompassing diverse programs and pedagogical contexts in anatomy and histology. ARanatomy was integrated through heterogeneous usage modalities, ranging from introductory practical sessions to supervised competency-focused activities.

At the University of Liège (Belgium), three cohorts (cohorts 1—3) used version 0 of the application. Cohort 1 included first-year medical and dental students (BA1, $n = 315$), and cohort 2 comprised first-year biomedical sciences students (BA1, $n = 189$), both using ARanatomy during introductory anatomy practicals. Cohort 3 consisted of second-year medical students (BA2, $n = 169$), who used the application in a supervised renal histology session. In this cohort, students were divided into two groups: the AM group completed a performance test before using the application, whereas the PM group completed it after use. Perception data were collected for all cohorts.

At the University of Lubumbashi (Democratic Republic of the Congo), cohort 4 included second-year medical students (BA2, $n = 1,200$) who used version 1 during introductory anatomy practicals. This version enabled partial usage instrumentation and the collection of perception, motivation, and pre/post test data.

Table 1 Evaluation questionnaire of ARanatomy application (adapted from WBLT).

Questionnaire Item
Learning
Working with ARanatomy helped me learn
The information provided by ARanatomy made my learning easier
The augmented reality (AR) component made my learning easier
Helped me learn the relationships between anatomical structures in 3D
Better visualization of the nephro-urinary system in space
The histological images included in the application helped me learn
The different magnification levels helped me learn
Design
The application interface was intuitive and easy to use
The 3D model was easy to manipulate (zoom and rotate)
The different features are easy to access
The screenshot and annotation features are an added value
Engagement
I found the application engaging
The application made learning fun
I would like to use the application again
I would show the application to my contacts

Evaluation questionnaire used to assess students' perceptions of the ARanatomy application, adapted from the Web-Based Learning Tools (WBLT) framework. The instrument covers three dimensions — learning, design, and engagement — and consists of items rated on a 4-point Likert scale ranging from strongly disagree to strongly agree.

At the Haute École de la Province de Liège (Belgium), cohort 5 consisted of first-year oral hygiene students (BA1, $n = 158$) who used version 2 during a supervised renal histology session. This fully instrumented version enabled detailed analysis of interaction behaviors alongside perception and motivation data.

Together, these cohorts represent substantial variability in institutional context, instructional design, and technological maturity.

Perception survey

Perception of the application was assessed using an adapted version of the Web-Based Learning Tools Framework [12], which evaluates three dimensions: learning (perceived learning support), design (usability and interface quality), and engagement (interest and user involvement). Items were tailored to ARanatomy's functionalities, focusing on 3D manipulation, spatial understanding, integration of histological content, and interface ergonomics. Lexical simplifications were applied for first-year cohorts to ensure item comprehension. All items were rated on a four-point Likert scale ranging from ''Strongly disagree'' (1) to ''Strongly agree'' (4) (Table 1).

Motivation assessment

Motivation was assessed in cohorts using instrumented versions of the application (cohorts 4 and 5) based on self-determination theory [7,8]. The questionnaire comprised 18 items distributed across four dimensions: intrinsic motivation, extrinsic motivation, amotivation, and positive emotions (Appendix 1 Table S1). Responses were collected on a four-point Likert scale, and dimension scores were summed to compute a global motivation index after reverse coding amotivation items. Motivation quartiles (Q1—Q4) were defined using empirical percentiles and used for subsequent analyses.

Study implementation and data availability

All cohorts contributed perception data. Usage analytics were available only for instrumented versions (cohorts 4 and 5). Pre-/post-test designs were implemented in two cohorts, aligned with their specific learning objectives: visuospatial ability assessments in cohort 4 and disciplinary histology performance in cohort 2. Analyses were therefore organized according to the type of available data within each cohort, rather than a single unified experimental design.

Statistical analyses

Statistical analyses were selected based on data distribution and variable type. Normality was assessed using the Shapiro-Wilk test. Descriptive statistics included means and standard deviations for normally distributed variables, medians and interquartile ranges for non-normal variables, and frequencies for categorical data.

Group comparisons employed parametric or non-parametric tests as appropriate. Associations between motivation, usage indicators, and performance were examined using Pearson or Spearman correlation coefficients, while relationships between motivational profiles and user categories were tested using Chi^2 analyses. Statistical significance was set at $P < 0.05$, and effect sizes were reported when relevant to support interpretation of the results.

Results

Perception of the tool

Perception data were collected across all five cohorts using the Web-Based Learning Tools Framework (WBLT), covering the learning, design, and engagement dimensions. This enabled cross-cohort comparisons while accounting for differences in pedagogical context and application version.

Learning dimension

Perceptions of learning potential varied markedly between cohorts (Fig. 3). Cohort 1 reported high satisfaction, with over 85% positive responses (''Agree'' or ''Strongly agree''). In contrast, cohort 2 showed a less favorable profile, with a predominance of negative responses. Cohort 3 displayed an

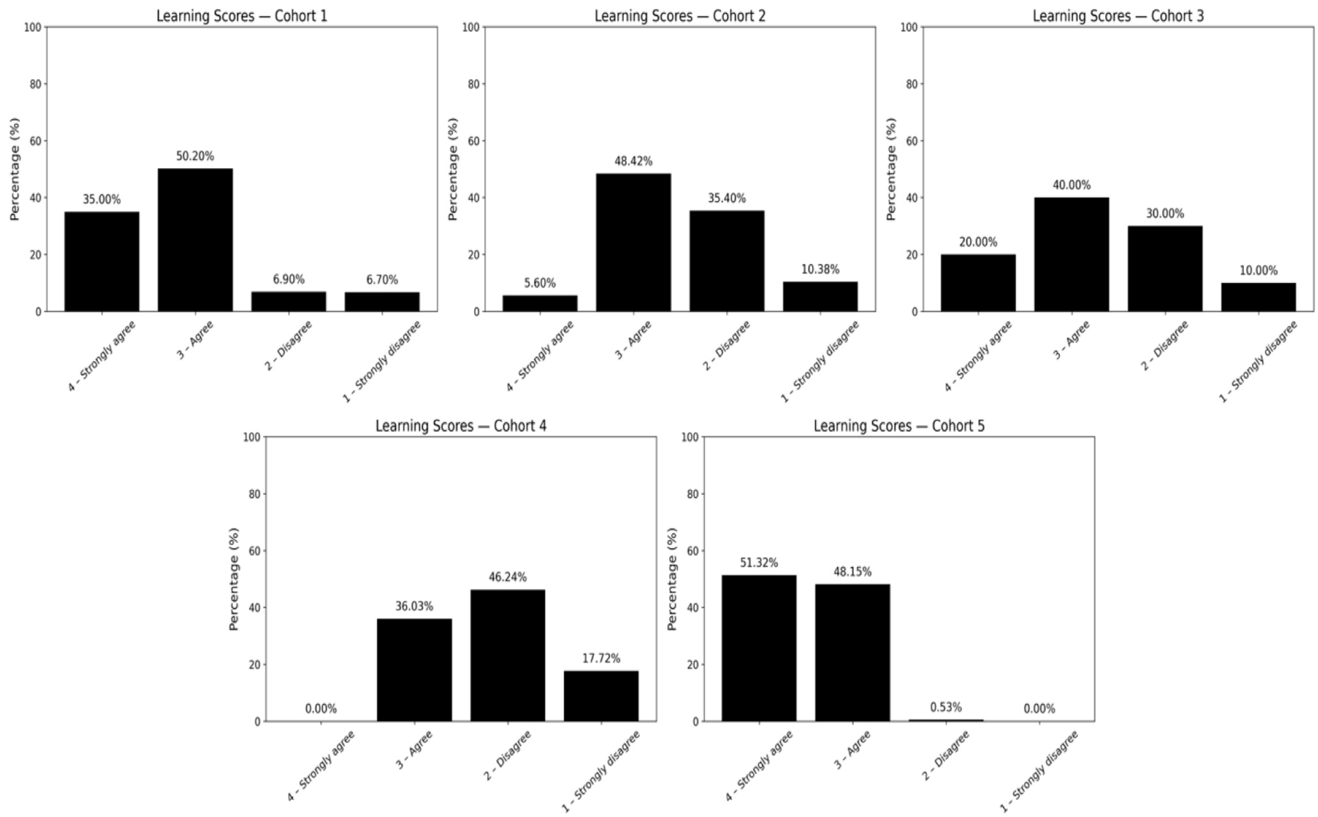

Figure 3 Distribution of responses related to learning scores across cohorts. Percentages represent levels of agreement on a 4-point Likert scale, highlighting marked inter-cohort variability in perceived learning value associated with contextual and implementation differences.

intermediate pattern, combining positive evaluations with substantial reservations.

Cohort 4 demonstrated the lowest perceived learning benefit, with a majority of ''Disagree'' or ''Strongly disagree'' responses. Conversely, cohort 5 showed near-unanimous endorsement, with more than 99% positive responses.

A Chi^2 test confirmed a highly significant association between cohort and perceived learning (χ^2 (12) = 757.1; P < 0.001; N = 2031), with a moderate-to-large effect size (Cramer's V = 0.353). Adjusted residuals revealed overrepresentation of highly positive responses in cohorts 1 and 5, and of negative responses in cohort 4, highlighting pronounced inter-cohort heterogeneity.

Design dimension

Perceptions of design quality also differed substantially across cohorts (Fig. 4). Cohorts 1 and 2 reported mostly positive evaluations of the interface and usability, whereas cohorts 3 and 4 expressed predominantly negative perceptions, particularly in cohort 4, where strong disagreement was prevalent. Cohort 5 displayed the most favorable design evaluations, with nearly 80% agreement, including a high proportion of ''Strongly agree'' responses.

These contrasts indicate a clear polarization between cohorts exposed to more ergonomically mature versions of the application and those encountering usability limitations.

Student engagement across cohorts

Engagement scores varied significantly across cohorts (Fig. 5). High perceived engagement was observed in cohorts 1, 2, and 5, whereas cohort 3 showed lower mean engagement and greater dispersion.

In cohorts 4 and 5, comparisons between ARanatomy users and non-users revealed no significant differences in engagement scores (cohort 4: P = 0.67; cohort 5: P = 0.61), although a slight descriptive advantage for users was observed in cohort 5. Overall, engagement appeared more strongly influenced by pedagogical context than by application use *per se*.

Performance outcomes associated with ARanatomy use

Academic performance in histology (cohort 2)

In the second-year medical cohort at ULiège (n = 122 tested students), academic performance in renal histology showed substantial interindividual variability (mean = 59.0% ± 21.3). Students assessed before or after using ARanatomy (AM vs. PM groups) achieved comparable scores (57.8% vs. 60.0%). The Mann-Whitney U-test revealed no statistically significant difference between groups, indicating no measurable effect of immediate ARanatomy exposure on short-term

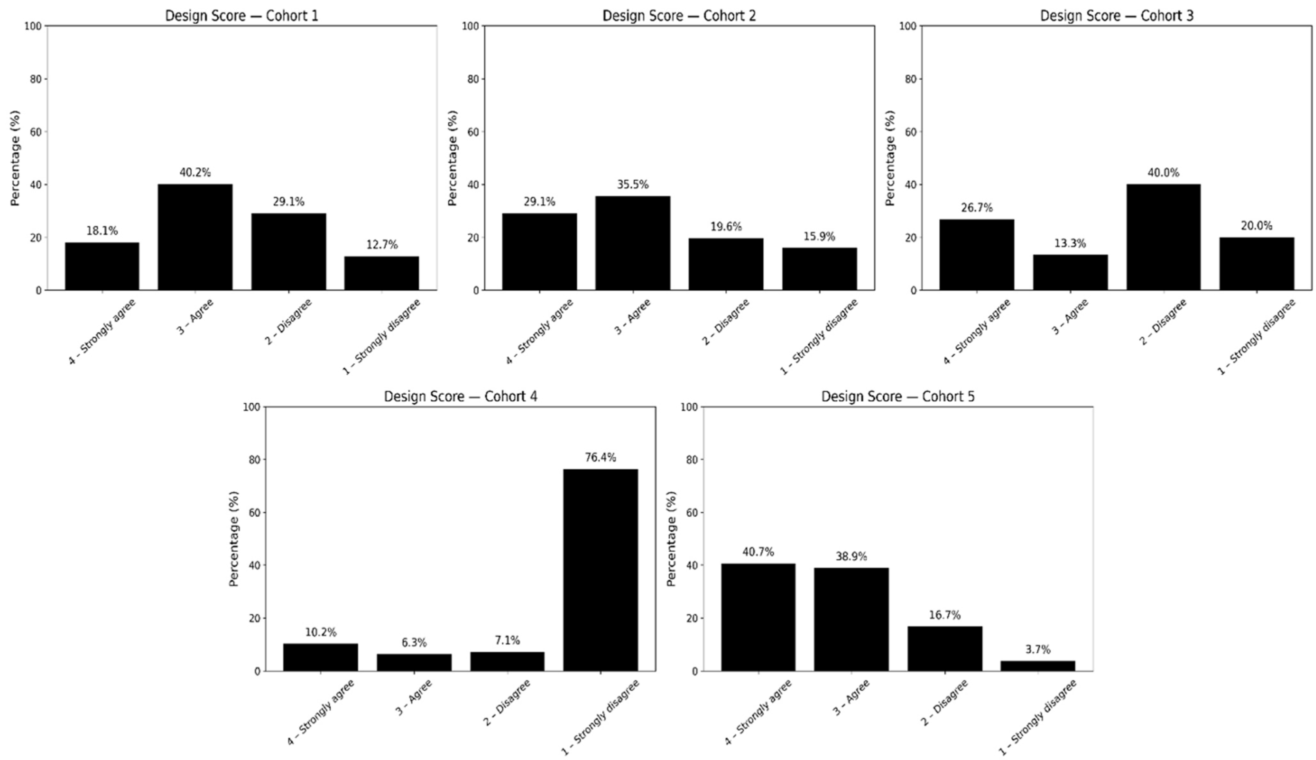

Figure 4 Distribution of agreement levels and perceptual profiles for the design dimensions across five evaluation cohorts. Percentages represent student responses on a 4-point Likert scale, illustrating contrasting perceptual profiles and marked differences in design appraisal between cohorts.

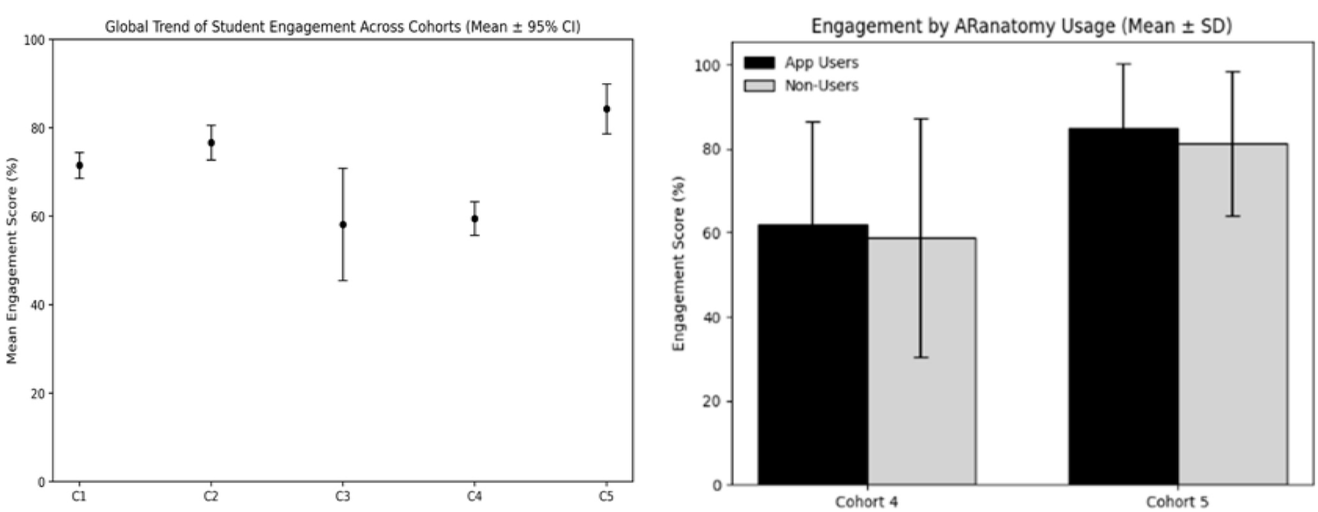

Figure 5 Student perception: overall trends across cohorts (C1—C5) and App/No App comparison in cohorts 4—5. Mean student engagement scores across the five cohorts (C1—C5), with 95% confidence intervals. The right panel presents mean engagement scores (±SD) for ARanatomy users and non-users in cohorts 4 and 5.

disciplinary performance, within the conditions of this assessment design (Fig. 6).

Visuospatial skills development (cohort 4)

Post-test performance revealed significantly higher visuospatial scores among ARanatomy users compared with non-users (41.4% vs. 31.7%; Welch *t*-test, $P < 0.05$) (Fig. 7). Usage intensity was positively correlated with post-test performance, particularly for the number of active days ($r = 0.30$; $P < 0.05$) and the number of views explored ($r = 0.29$; $P < 0.05$). Other indicators, such as session duration, showed positive but non-significant trends, suggesting a dose-response relationship between usage and performance.

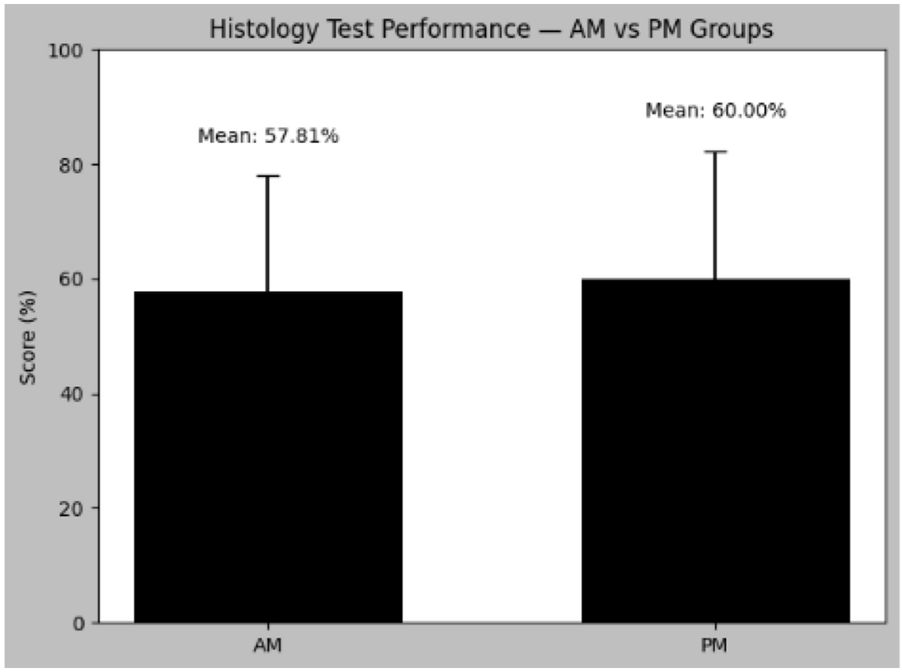

Figure 6 Cohort 2: distribution of histology test scores in AM and PM student groups. Bars represent mean test scores expressed as percentages for each group. Error bars indicate score dispersion around the mean. The figure summarizes the central tendency and variability of performance according to the assessment scheduling.

Pre-/post-test comparison (cohort 4)
Among students completing both tests ($n = 103$), descriptive trends suggested performance gains among users and a decline among non-users. However, paired t-tests showed no significant pre-/post-test differences within either group, nor significant differences in score changes between users and non-users. These results indicate that observed performance differences cannot be conclusively attributed to application use.

Motivational factors and application use (cohorts 4 and 5)
Overall motivation levels were high in both cohorts, with higher mean motivation observed in cohort 5. Users displayed slightly higher motivation scores than non-users, although these differences were descriptive only.

Analysis by motivation quartiles revealed a higher proportion of highly motivated students (Q4) among users compared with non-users in both cohorts. Intrinsic motivation and positive emotions were positively associated with usage intensity, whereas amotivation showed a consistent negative relationship with application use. In cohort 5, extrinsic motivation and positive emotions were strongly correlated with time spent in AR mode.

These patterns suggest that motivation is closely linked to actual engagement with the application, particularly its immersive features, and may act as a key modulator of usage behaviors.

Discussion

This multicentric study examined the educational impact of ARanatomy across heterogeneous pedagogical contexts by combining perception data, motivation profiles, instrumented usage traces, and performance outcomes. Overall, the results show that the effects of ARanatomy are neither uniform nor systematic. Rather, they depend strongly on (1) the pedagogical context of implementation, (2) students' motivation and actual intensity of use, and (3) the type of competence assessed. Beyond a tool-centered perspective, ARanatomy should be understood as a flexible and instrumented learning environment whose effects emerge from the interaction between technology, context, and learner characteristics.

The findings partially support H1 and H2 and conditionally support H3. The use of ARanatomy was not associated with higher perceived engagement *per se* (H1), but intensive and diversified use was associated with higher visuospatial performance and differentiated motivational profiles (H2). Finally, the effects on performance varied depending on pedagogical context and outcome type, with clear benefits for visuospatial skills but no measurable effect on immediate disciplinary performance (H3). This pattern suggests that AR technologies do not act as universal performance enhancers but as context-sensitive mediators of learning processes.

The primacy of pedagogical context (macro-level)

Marked differences in perception and engagement across cohorts highlight the central role of pedagogical context. Cohorts exposed to stable, ergonomically mature versions of the application within well-structured instructional settings reported highly positive evaluations, whereas less structured contexts using earlier versions yielded critical perceptions. These results confirm that immersive technologies cannot be evaluated independently of their pedagogical context. Instead, their educational value emerges through

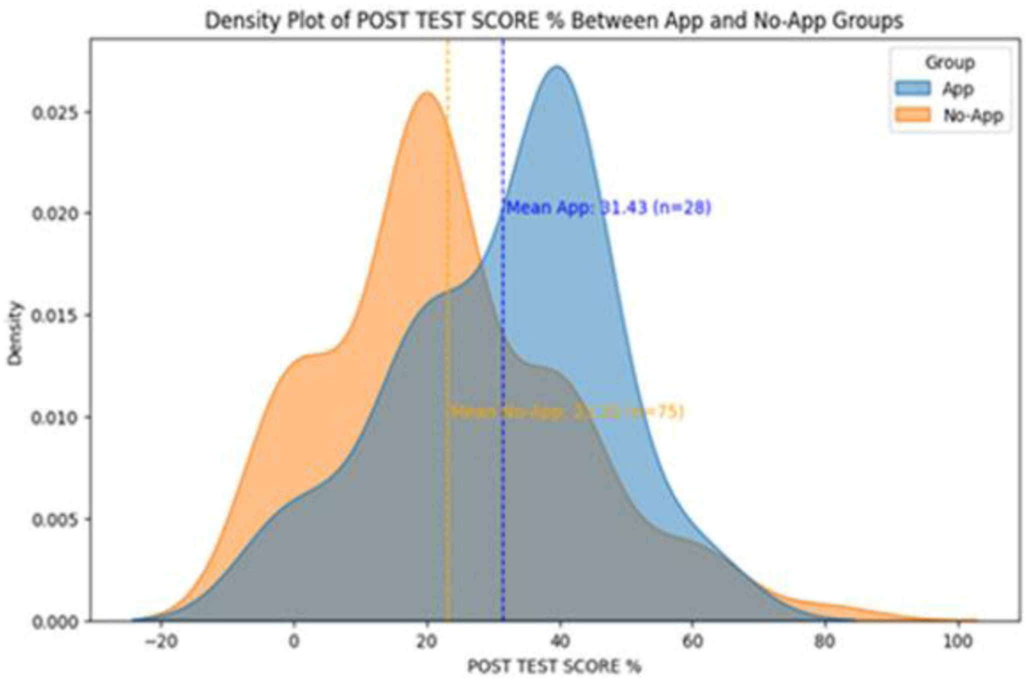

Figure 7 Distribution of post-test score by ARanatomy usage. Density curves represent score distributions for students who used the application (App) and those who did not (No-App). Vertical dashed lines indicate the mean post-test score for each group. The distribution for users is shifted toward higher scores compared with non-users.

alignment with instructional objectives, guidance, and curricular integration, as consistently reported in the literature [1,2].

Importantly, engagement levels appeared more closely tied to the broader learning environment than to the mere availability of AR. This aligns with evidence suggesting that immersive technologies act as amplifiers of existing pedagogical conditions rather than independent drivers of engagement.

Crucially, the variability observed across cohorts reflects not a limitation of the tool, but its pedagogical plasticity. ARanatomy adapts to diverse educational ecosystems, from resource-constrained contexts such as Lubumbashi — where it may compensate for limited access to physical materials — to well-equipped environments such as Liège, where it functions as a complementary and enrichment tool. This highlights its capacity to support differentiated instructional strategies rather than imposing a standardized model.

A key strength of this study lies in its multicentric design, which enables a structured comparison of contrasted educational contexts and their influence on perceptions, motivation, and usage patterns in augmented reality learning environments. The results do not reflect random variability across cohorts, but rather distinct and coherent profiles shaped by institutional context.

In the Belgian university cohorts (ULiège, C1—C3), results reveal marked heterogeneity in student perceptions, particularly in the learning and design dimensions of the WBLT framework. Profiles range from high levels of endorsement to more critical evaluations, suggesting that variations in pedagogical integration — from introductory exposure to more structured activities — combined with higher academic expectations, contribute to differentiated experiences of the tool.

In contrast, the non-university higher education cohort (HEPL, C5) is characterized by highly positive and homogeneous perceptions across all WBLT dimensions, associated with elevated levels of intrinsic motivation and positive emotions. This profile suggests a strong alignment between the tool, pedagogical objectives, and supervision modalities, within a professionally oriented training context where digital tools are closely linked to future practice.

The cohort from the University of Lubumbashi (C4) presents a distinct configuration, characterized by more critical perceptions in the Learning dimension, yet accompanied by significant levels of engagement and usage. This apparent discrepancy between reported perceptions and actual usage suggests specific appropriation dynamics, in which digital tools may simultaneously represent both an opportunity for access and a challenge in terms of familiarity, usage conditions, and pedagogical mediation.

The combined analysis of WBLT perceptions, motivational profiles, and instrumented usage data thus reveals differentiated configurations across contexts: heterogeneity and high expectations in Belgian university settings, strong alignment and adherence in non-university higher education, and pragmatic appropriation in the Sub-Saharan African context. These findings indicate that augmented reality learning environments do not produce uniform effects, but rather act as amplifiers of the pedagogical and cultural conditions in which they are implemented.

This interpretation is consistent with prior research showing that the effectiveness of educational technologies depends strongly on implementation conditions and pedagogical alignment [13,14], particularly in heterogeneous or resource-constrained contexts [15]. The multicentric nature of this study therefore constitutes not only a methodological strength but also an analytical framework for understanding contextual dynamics in digital learning environments.

Usage intensity and motivation as key modulators (meso-level)

While simple comparisons between users and non-users yielded limited differences, instrumented analyses revealed that depth and regularity of use represent critical differentiators. Intensive users demonstrated higher visuospatial performance and more favorable motivational profiles. Intrinsic motivation and positive emotions were consistently associated with richer exploration of the application, whereas amotivation was associated with minimal use.

These findings are in line with Self-Determination Theory [7,8], which posits that intrinsic motivation fosters autonomous and sustained engagement. In this regard, ARanatomy's interactive and spatial affordances appear to support learning only when students actively invest in exploration rather than passively interact with the tool. Thus, effectiveness depends on the quality of engagement rather than exposure alone. This shifts the interpretation from access to appropriation: the presence of the tool is insufficient, and its educational value is contingent upon active, sustained, and meaningful use. ARanatomy therefore operates as an affordance-rich environment whose potential is actualized through learner engagement.

Differential effects on learning outcomes (micro-level)

The study highlights differing patterns between disciplinary performance and visuospatial skills. Immediate academic performance in histology was not influenced by ARanatomy use, particularly in contexts involving short, one-time exposure and highly structured assessments. In contrast, visuospatial abilities — closely aligned with the cognitive affordances of AR — were higher among users, with stronger effects observed as usage intensity increased.

This pattern aligns with previous research showing that 3D and AR environments preferentially support spatial cognition rather than declarative knowledge acquisition [3,4]. This suggests that AR may be more effective when the targeted competence directly matches the underlying cognitive processes supported by the technology. These findings reinforce the importance of constructive alignment between technological affordances and targeted learning outcomes, positioning ARanatomy as particularly relevant for tasks involving spatial reasoning and mental manipulation of anatomical structures.

Selection effects and limits of causal inference

Although users outperformed non-users in visuospatial tasks, pre-test differences and the absence of significant pre-/post-test effects prevent causal attribution. Students with stronger initial visuospatial abilities may also be more inclined to adopt immersive environments, resulting in a selection effect. Thus, these associations should not be interpreted as causal.

This highlights the need to account for learner characteristics, such as digital fluency and perceived competence, when evaluating immersive technologies. Future studies should adopt designs explicitly addressing self-selection mechanisms and longitudinal learning trajectories.

Methodological contributions and practical implications

This study highlights the value of combining multicentric designs with learning analytics data. Instrumented usage traces enable objective identification of learner profiles and reveal engagement mechanisms that remain invisible in self-reported data alone [9,10].

Practically, ARanatomy should be integrated into structured pedagogical strategies promoting sustained and guided use. Occasional or unguided exposure is unlikely to yield measurable learning gains, whereas structured integration targeting spatial reasoning may leverage the technology's full educational potential. More broadly, these findings support a shift from technology adoption to pedagogical orchestration, where immersive tools are embedded within intentional learning designs tailored to specific contexts and learner needs.

Conclusion

ARanatomy functions not as a universal learning enhancer but as an instrumented learning environment whose impact depends on context, motivation, and usage intensity. Its strongest contribution lies in supporting visuospatial cognition, while effects on immediate disciplinary performance remain conditional and context-dependent. Its added value lies in its flexibility and adaptability across diverse educational settings, enabling context-sensitive implementations ranging from compensatory to enrichment functions. These findings call for pedagogically aligned, analytics-informed deployment of AR technologies in anatomy and histology education.

Use of artificial intelligence

The authors used artificial intelligence tools to assist with language editing, English translation, and formatting of the manuscript. These tools were not used to generate scientific content, interpret data, or draw conclusions. All scientific aspects of the work were developed and validated by the authors.

Ethics statement

These datasets were collected and processed in accordance with the Humanities and Social Sciences Ethics Committee (Ref. 2023-18), and in compliance with the European General Data Protection Regulation.

Authors' contribution

Dr. David Mutombo Mwembo: conceptualization, investigation, writing — original draft, methodology, validation, visualization, writing — review & editing, project administration.

Dr. Allyson Fries: investigation, writing — original, draft, methodology, validation, visualization, writing — review & editing, supervision.

Prof. Pierre Bonnet: funding acquisition, writing — review & editing, project administration, supervision.

Dr. Arsène Sul Ilunga Sul: investigation, methodology, writing — review & editing, formal analysis.

Mr. Olivier Prygiel: investigation, methodology, writing — review & editing, formal analysis.

Prof. Aude Lagier: funding acquisition, writing — review & editing, project administration, supervision.

Prof. Willy Arung Kalau: funding acquisition, project administration, supervision.

Prof. Stéphane Sobczak: funding acquisition, writing — review & editing, supervision.

Prof. Michael Schyns: conceptualization, investigation, funding acquisition, methodology, visualization, formal analysis, supervision.

Prof. Valérie Defaweux: corresponding author, submitting author, conceptualization, investigation, funding acquisition, writing — original draft, methodology, validation, visualization, writing — review & editing, formal analysis, project administration, supervision, resources.

All authors have read and approved the final manuscript.

Funding

This work was supported by the Anatomy Laboratory of the University of Liège [ARES Project-PFS-2022-RDC-VanSintJan: University of Lubumbashi, University of Liège, and Université Libre de Bruxelles]; by The Université du Québec à Trois-Rivières [FIP (*fonds d'innovation pédagogique)*]; and by the University of Liège [*fonds de soutien à l'enseignement ULiège*].

Disclosure of interest

The authors declare that they have no competing interest.

Acknowledgments

The authors would like to thank Mrs. Marie Goffin for her valuable assistance in data collection and statistical analysis and Mr. Alexis Jacquemin (Michael Shyns' team) for his contributions to the application's graphic design and visual representation of anatomically relevant structures in virtual reality.

Appendix 1. Table S1 Motivation survey items categorized by motivational pillar

Item	Question	Motivation pillar
Q1	For me, it is important to be familiar with digital learning environments dedicated to anatomy education because it allows me to stay up to date with modern developments	Intrinsic motivation
Q2	The use of digital tools in anatomy education creates a positive working atmosphere	Positive emotions
Q3	It is important for me to master digital learning environments in anatomy because this will help me become a better physician	Extrinsic motivation
Q4	Using digital tools in anatomy teaching makes the course more enjoyable	Positive emotions
Q5	If I had the choice, I would not use digital environments to learn anatomy	Amotivation
Q6	When using digital tools to study anatomy, making mistakes is normal	Positive emotions
Q7	It is important for me to master digital anatomy learning environments because they help me acquire knowledge more easily	Intrinsic motivation
Q8	I enjoy using digital environments when learning anatomy	Positive emotions
Q9	I never feel bored when using digital anatomy learning environments	Positive emotions
Q10	By using digital environments for learning anatomy, I have learned interesting things	Positive emotions
Q11	It is important for me to be familiar with digital anatomy learning environments because they facilitate my learning in other subjects of the curriculum	Intrinsic motivation
Q12	During class, I feel proud of my achievements	Positive emotions

Q13	In my academic journey, we laugh a lot	Positive emotions
Q14	It is important for me to master digital anatomy learning environments because it will make me more competitive in the job market	Extrinsic motivation
Q15	It is cool to use digital anatomy learning environments	Positive emotions
Q16	It is important for me to master digital learning environments in order to become a reference person in my university and among other students	Extrinsic motivation
Q17	I think that using digital anatomy learning environments is a complete waste of time	Amotivation
Q18	I feel comfortable using digital anatomy learning environments	Positive emotions

Items Q5 and Q17 (Amotivation) were reverse scored in the calculation of the global motivation index. Intrinsic and extrinsic motivation items were adapted to reflect the specific context of digital anatomy learning. The questionnaire was administered in French and translated into English for reporting purposes.

References

[1] 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. Anatom Sci Educ 2024;17(3):514–28, http://dx.doi.org/10.1002/ase.2397.

[2] Tene T, Bonilla García N, Coello-Fiallos D, Borja M, Vacacela Gomez C. A systematic review of immersive educational technologies in medical physics and radiation physics. Front Med 2024;11:1384799, http://dx.doi.org/10.3389/fmed.2024.1384799.

[3] Bogomolova K, Hierck BP, Looijen AEM, Pilon JNM, Putter H, Wainman B, et al. Stereoscopic three-dimensional visualisation technology in anatomy learning: a meta-analysis. Med Educ 2021;55(3):317–27, http://dx.doi.org/10.1111/medu.14352.

[4] Koh MY, Tan GJS, Mogali SR. Spatial ability and 3D model colour-coding affect anatomy performance: a cross-sectional and randomized trial. Sci Rep 2023;13:7879, http://dx.doi.org/10.1038/s41598-023-35046-2.

[5] Masoumian Hosseini ST, Qayumi K, Pourabbasi A, Haghighi E, Sabet B, Koohpaei A, et al. Are we ready to integrate modern technologies into the medical curriculum? A systematic review. Discov Educ 2025;4(1):114, http://dx.doi.org/10.1007/s44217-025-00521-7.

[6] Tzortzoglou F, Sofos A. Evaluating the usability of mobile-based augmented reality applications for education: a systematic review. In: Bratitsis T, editor. Research on e-learning and ICT in education. Springer; 2023. p. 105–35, http://dx.doi.org/10.1007/978-3-031-34291-2_7.

[7] Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68–78, http://dx.doi.org/10.1037/0003-066X.55.1.68.

[8] Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020;61:101860, http://dx.doi.org/10.1016/j.cedpsych.2020.101860.

[9] Sakr A, Abdullah T. Virtual, augmented reality and learning analytics impact on learners and educators: a systematic review. Educ Inf Technol 2024;29(15):19913–62, http://dx.doi.org/10.1007/s10639-024-12602-5.

[10] Singh M, Bangay S, Sajjanhar A. Augmented reality enhanced analytics for education: a systematic review. J Learn Anal 2025;12(3):126–55, http://dx.doi.org/10.18608/jla.2025.8447.

[11] Microsoft Corporation. Microsoft power BI desktop [computer software]; 2024 [https://powerbi.microsoft.com/].

[12] Kay R. Evaluating learning, design, and engagement in web-based learning tools (WBLTs): the WBLT Evaluation Scale. Comput Hum Behav 2011;27(5):1849–56, http://dx.doi.org/10.1016/j.chb.2011.04.007.

[13] Kirkwood A, Price L. Technology-enhanced learning and teaching in higher education: what is ''enhanced'' and how do we know? A critical literature review. Learn Media Technol 2014;39(1):6–36, http://dx.doi.org/10.1080/17439884.2013.770404.

[14] Selwyn N. Education and technology: key issues and debates, 2nd ed. Bloomsbury Academic; 2017.

[15] Rodriguez-Segura D. EdTech in developing countries: a review of the evidence. World Bank Res Observ 2021;36(2):171–203, http://dx.doi.org/10.1093/wbro/lkab011.