Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transitioning from Passive to Active Learning: An Assessment of Game-Based Pedagogy in Thoracic Anatomy Education at the University of Rwanda

Victor Bassey Archibong

victorarchibong3@gmail.com

University of Rwanda

Prakash Manickam Kumarasamy

Newcastle University Medicine

Fentahun Adane

University of Rwanda

Hussen Abdu

University of Rwanda

Samson Olorunnado

University of Rwanda

Usman Ibe

University of Rwanda

Abdullahi Mohammed

University of Rwanda

Catherine Wali

University of Rwanda

David Chibuike Ikwuka

University of Rwanda

Ekom Monday Etukudo

National Defence University

Julien Gashegu

University of Rwanda

Research Article

Keywords: Anatomy education, Gamification, Simulation-based learning, Thoracic anatomy, Medical education, Active learning, Low-resource settings

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11241218/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Anatomy teaching often emphasises memorisation over clinical understanding. Thoracic surface anatomy is particularly challenging because it requires three-dimensional spatial reasoning and integration with palpation and auscultation. Evidence on low-cost, wearable mannequin-based gamification in resource-limited settings is limited.

Methods

We conducted a quasi-experimental, parallel-group, post-test-only educational study among 133 preclinical medical and dental students at the University of Rwanda. On the lecture day, participants were randomly allocated to eight groups: one control group (A; *n* = 21) and seven intervention groups (B–H; *n* = 16 each). They were taught using gamified, role-based learning with a wearable thoracic mannequin apron. Outcomes were four-item gamified quiz performance and four five-point Likert measures of perceived understanding, heart-sound learning, lung-landmark learning, and fun/engagement. Data were analysed using MANOVA, one-way ANOVA with Welch's test, Games–Howell post-hoc comparisons, and Mann–Whitney *U* test.

Results

Quiz accuracy was high (86–99%). All groups achieved full marks on three of four items; on the remaining item, Groups B–F scored 100%, while Groups A, G, and H scored 75%. Perceived learning was favourable (composite mean 4.40, SD 0.80). The multivariate group effect was not significant [Pillai's trace = 0.252, $F(28, 500) = 1.201$, $p = 0.222$]. Fun/engagement showed a significant omnibus effect [$F(7, 125) = 2.189$, $p = 0.039$, partial $\eta^2 = 0.109$], but no pairwise comparison was significant. Perceived-learning scores did not differ significantly between groups ($p = 0.177$).

Conclusion

Gamified mannequin-based teaching was well received and associated with high short-term performance but did not demonstrate superiority over conventional model-based teaching. Ceiling effects, small groups, and the post-test-only design limit inference. Larger controlled studies with baseline testing, participant-level objective outcomes, and longer follow-up are warranted.

INTRODUCTION

Anatomy education is a fundamental component of medical training because it provides the structural basis for understanding physical examination, diagnostic reasoning, imaging, and clinical procedures. A

sound knowledge of anatomy is also essential for identifying and palpating clinically important surface landmarks. Although anatomy has traditionally been taught through lectures, cadaveric material, specimens, and anatomical models, contemporary medical education increasingly emphasises approaches that actively engage learners and promote the application rather than simple memorisation of anatomical knowledge [1, 2]. More broadly, evidence from higher education indicates that active-learning approaches can improve student performance compared with traditional lecture-dominated instruction [3, 4].

Surface anatomy is particularly important because it connects internal anatomical structures with features identifiable on the living body and provides a foundation for physical examination and clinical procedures. Students must integrate three-dimensional spatial relationships with inspection, palpation, percussion, and auscultation, making surface anatomy an important bridge between preclinical anatomical knowledge and clinical competence [5, 6]. Clinically integrated anatomy teaching has therefore been advocated as a means of improving students' ability to apply anatomical knowledge in clinical settings[2, 7] .

Thoracic surface anatomy is especially relevant to clinical examination because accurate identification of the sternum, sternal angle, ribs, intercostal spaces, midclavicular lines, cardiac valve auscultation areas, and pulmonary surface landmarks is required during examination of the cardiovascular and respiratory systems. Conventional static models can demonstrate these structures, but they do not fully reproduce the experience of identifying landmarks on a living body. Living-anatomy approaches, including peer examination and the use of living models, can provide opportunities for students to relate underlying structures to palpable surface landmarks and clinical examination techniques [8]. Recent work-based surface-anatomy teaching also supports contextualised, active and experiential instruction for developing cognitive, psychomotor, and affective competencies relevant to clinical practice [5].

Game-based learning and gamification have emerged as potentially useful strategies for increasing active participation and engagement in health professions education. Gamification involves incorporating game attributes such as challenge, interaction, immediate feedback, and structured participation into otherwise non-game educational activities [9]. Systematic reviews suggest that gamified and serious-game approaches can increase learner engagement and motivation and may improve some learning outcomes, although the magnitude of educational benefit varies according to intervention design, educational context, and outcome measure [9–11]. Simulation-based education similarly provides learners with structured opportunities to practise and integrate knowledge and skills, with evidence supporting its value across health professions education [12].

An approach particularly relevant to thoracic surface anatomy is the use of a transparent wearable apron in which one student acts as a human mannequin while peers identify, palpate, and trace anatomical landmarks. Kumarasamy[13] described such an approach for teaching chest-wall anatomy and clinical skills: students worked in groups of three, with one student wearing the transparent apron, another identifying anatomical landmarks, and a third tracing the landmarks before percussion and auscultation

were practised. This format enables learners to relate anatomical structures directly to the living body while engaging in peer-assisted, hands-on learning. It is also consistent with experiential-learning principles, in which knowledge is developed and consolidated through participation in authentic or contextually meaningful experiences followed by interaction and reflection [14].

At the School of Medicine and Pharmacy, College of Medicine and Health Sciences, University of Rwanda, anatomy education incorporates conventional resources alongside newer three-dimensional technologies. The Anatomage Table, for example, has been introduced as an additional anatomy-learning resource, and University of Rwanda students have reported favourable perceptions of its contribution to visualisation, understanding of anatomical relationships, interest, and enjoyment of anatomy learning [15]. Nevertheless, complementary low-cost approaches that permit direct identification and palpation of surface landmarks may provide an additional means of strengthening the link between anatomical knowledge and clinical examination.

Although gamification, simulation, living anatomy, and technology-enhanced anatomy education have received increasing attention, controlled evidence specifically evaluating wearable apron-based human-mannequin instruction for thoracic surface anatomy remains limited. The previously reported human-mannequin approach demonstrated its practical feasibility and received favourable anecdotal feedback, but it did not provide controlled comparative evidence of educational effectiveness [13]. Consequently, further evaluation is required to determine whether this approach provides measurable educational advantages over conventional anatomical models.

The present study therefore employed a controlled educational intervention to evaluate a gamified, wearable human-mannequin approach for teaching thoracic wall surface anatomy among preclinical medical and dental students. We hypothesised that the interactive approach would promote student engagement and understanding of thoracic surface anatomy and clinically relevant landmarks compared with conventional plastic-model instruction. The study was intended to determine whether mannequin-assisted, game-based pedagogy could serve as a feasible supplementary strategy for anatomy education and provide evidence to inform the development of active and clinically integrated anatomy curricula.

METHODS

Study setting

The study was conducted at the School of Medicine and Pharmacy (SoMP), College of Medicine and Health Sciences (CMHS), University of Rwanda (UR), Huye Campus. Huye District is one of the districts of Rwanda's Southern Province [16].

Study design

This study employed a quasi-experimental controlled educational intervention design to evaluate the effectiveness of a gamified, mannequin-based approach to teaching thoracic wall anatomy compared with traditional plastic anatomical models.

Participants

The study was conducted among undergraduate preclinical one (PCL 1) students enrolled in the thoracic wall anatomy module at SoMP, CMHS, UR, Huye campus. All participants had previously completed foundational gross anatomy coursework (Limbs and back anatomy) but had not yet received formal instruction on thoracic wall surface anatomy and clinical correlation. Participation was voluntary, and informed consent was obtained before inclusion in the study.

Criteria for inclusion and exclusion

Inclusion criteria: Year one medical and dental surgery students of UR who were at the Huye campus UR, and consented to the study were included.

Exclusion criteria: Year one medical and dental surgery students in other campuses of UR were excluded. Year one medical and dental surgery students who did not consent were also excluded.

Sample size

The source population consisted of 170 first-year medical and dental surgery students. The minimum sample size was determined using the Krejcie and Morgan sample-size determination table [17], an established approach that has continued to be applied in more recent studies, including research conducted in Uganda [18]. Based on a target population of 170, the table indicated a minimum required sample of 118 participants. A total of 133 students were ultimately recruited based on their availability for the scheduled teaching session and willingness to participate.

Pre-intervention preparation

Seven days before the scheduled session, all participants received standardised preparatory materials: lecture notes covering the thoracic wall and thoracic viscera, and pre-recorded instructional videos explaining anatomical landmarks, surface anatomy and clinical relevance. Students were asked to review these materials in advance so that the session focused on application and reinforcement rather than on first exposure to the content.

Group allocation

On the day of the lecture, participants were randomly selected into eight groups (groups A – H). Group A served as the control (n = 21). The seven intervention groups (groups B to H) consisted of 16 members each (n = 16 x 7). Therefore, the total number of participants was 133 (N = 133). The intervention groups used a wearable human-mannequin transparent-apron model developed at Newcastle University Medicine, Malaysia (copyright no. CRLY2024J06464) and previously described as an approach to teaching chest-wall surface anatomy and clinical skills [13]. The control group used a conventional plastic thoracic wall model. All groups received the same instructional content objectives and equal teaching duration.

Intervention: gamified mannequin-based learning

In the intervention groups, learning was conducted using a wearable human mannequin apron model. Within each group, students rotated through three active roles:

1. Mannequin Role: One student wore the thoracic anatomy apron model.
2. Landmark Identifier: One student identified anatomical landmarks on the chest wall using palpation.
3. Tracer: One student traced identified landmarks with a marker on the apron.

Clinical palpation techniques were performed directly on the student wearing the mannequin apron to determine the Sternal angle, sternal line, intercostal spaces, midclavicular line, auscultation points of the valves of the heart (aortic, pulmonary, tricuspid, and mitral), and lung auscultation points (superior, middle, and inferior lobes). The session lasted three hours, during which students also practiced simulated auscultation of heart and lung sounds on the mannequin model to reinforce clinical correlation. The approach was consistent with experiential and living-anatomy teaching principles, in which learners apply anatomical knowledge directly to the human body in a supervised educational environment [8, 14].

Control group: Conventional model-based learning

The control group received the same content objectives and instructional time (three hours), but learning was conducted using a standard plastic thoracic wall model. Students identified and discussed anatomical landmarks on the static model under instructor guidance, without role-playing.

Data collection method

Data were collected using a Google Forms-based questionnaire. The questionnaires were developed by the study investigators specifically for this study, informed by the learning objectives of the topic [Supplementary file]. The questionnaire consisted of three sections; The first section collected participants' sociodemographic information, including sex and program of study. The second section involved a gamified formative assessment consisting of multiple-choice and applied clinical questions on surface anatomical landmarks, thoracic wall structures, and auscultation points. Each student was provided with three color-coded response cards (blue, yellow, and red) and responded to each question

by raising the card corresponding to their selected answer. The answer keys were color-coded to match the response cards, facilitating immediate feedback and encouraging group engagement. The quiz was designed to assess short-term knowledge acquisition and reinforce learning through interactive participation. Part three assessed the students' opinion on how the human mannequin model (independent variable) improved understanding of landmarks, heart and lung sound auscultation, and learning. Relevant questionnaire concepts were informed by the student-perception approach reported by Archibong et al. [15] in their evaluation of anatomy-learning technology at the University of Rwanda. All 133 questionnaires were duly filled out by the participants (100% response rate).

Outcome measures

Two outcome domains were evaluated. Objective short-term knowledge acquisition was assessed with a four-item gamified formative quiz addressing the same four landmark topics in the same order: Question 1 (Q1), the vertebral level of the sternal angle; Question 2 (Q2), the event not occurring at the sternal angle; Question 3 (Q3), the aortic valve auscultation site; and Question 4 (Q4), the right middle lobe auscultation site. This quiz was scored independently in two ways. Individual-level responses from all 133 participants were summarised as the percentage selecting each option. Separately, performance was also scored at the group level, with each of the four items worth 25 points towards a maximum group score of 100. These group scores were assigned directly, not calculated from the individual percentages. Perceived learning was assessed with four five-point Likert items scored from 1 (strongly disagree) to 5 (strongly agree): Q1, perceived understanding; Q2, perceived learning of heart sounds; Q3, perceived understanding of lung auscultation landmarks; and Q4, fun/engagement with learning. Higher scores indicated more favourable perceptions. A composite perceived-learning score was calculated for each participant as the mean of Q1–Q4.

Statistical analysis

Data were analysed using IBM SPSS Statistics version 27. Sociodemographic characteristics were summarised using frequencies and percentages, while perceived-learning outcomes were presented as means and standard deviations. A composite perceived-learning score was calculated as the mean of Q1–Q4.

Differences among the eight study groups across the four perceived-learning domains were assessed using one-way MANOVA. Because Box's M test indicated unequal covariance matrices, Pillai's trace was used as the primary multivariate statistic. Univariate ANOVA was conducted for each outcome, with Welch and Brown–Forsythe robust tests applied to Q3 and Q4 where Levene's test indicated unequal variances. Significant pairwise differences were explored using Games–Howell tests, with Tukey HSD and Bonferroni-adjusted comparisons examined as supplementary analyses. A Bonferroni-adjusted significance level of $p < 0.0125$ was applied across the four domain-specific omnibus tests. Partial eta squared (partial η^2) was reported for multivariate and univariate effect sizes.

Composite perceived-learning scores were compared across groups using one-way ANOVA and, where variance homogeneity was violated, Welch's ANOVA with Games–Howell post-hoc tests. Eta squared

(η^2) was reported as the effect size. An exploratory Mann–Whitney U test compared composite scores between Medicine and Dental Surgery students at a two-sided significance level of $p < 0.05$.

RESULTS

Participant characteristics

A total of 133 students participated in the study and provided complete demographic information. Of these, 85 (63.9%) were male, and 48 (36.1%) were female (Fig. 1A). Most participants were enrolled in Medicine (n = 108, 81.2%), while 25 (18.8%) were Dental Surgery students (Fig. 1B). Participants were distributed across eight study groups, comprising one control group (Group A; n = 21, 15.8%) and seven intervention groups (Groups B–H; n = 16, 12.0% each) (Table 1).

(A) Sex; (B) programme of study.

Table 1
Descriptive statistics of the study groups

Characteristic	***N***	**%**
Group A (control)	21	15.79
Group B	16	12.03
Group C	16	12.03
Group D	16	12.03
Group E	16	12.03
Group F	16	12.03
Group G	16	12.03
Group H	16	12.03
Total	133	

Percentages are of the total sample (*N* = 133). Groups B–H were the seven intervention groups. % were rounded to whole.

Individual responses to the gamified quiz

Participants demonstrated high accuracy across the four quiz questions assessing anatomical knowledge and its clinical applications (Table 2). Most correctly identified the vertebral level of the sternal angle (97%) and recognised that looping of the right recurrent laryngeal nerve is not an anatomical event associated with the sternal angle (98%).

Performance was similarly high on the clinically oriented questions. The correct aortic valve auscultation site was identified by 86% of participants, while 99% correctly identified the auscultation region of the right middle lobe of the lung. The aortic valve auscultation question had the lowest percentage of correct responses, with 11% selecting the left parasternal location and 3% selecting the right mid-clavicular location.

Table 2
Participants' responses to the gamified quiz

1	Response option	Responses (%)
	The sternal angle lies at	
	T1–T2	1
	T3–T4	2
	T4–T5	**97**
2	**This is not an anatomical event at the sternal angle**	
	Counting of ribs	1
	Division of the mediastinum into superior and inferior	1
	Looping of the right recurrent laryngeal nerve	**98**
3	**Aortic valve sound auscultation is at**	
	2nd IC space on the left parasternal line	11
	2nd IC space on the right sternal line	**86**
	2nd IC space on the right mid-clavicular line	3
4	**Middle lobe of the right lung auscultation**	
	Auscultate below the 6th rib	1
	Auscultate between the 4th and 6th ribs	**99**
	Auscultate between the 1st and 4th ribs	0

The questions 1 to 4 are italicized. The correct response to each item is shown in bold. IC, intercostal. (% of respondents, *N* = 133).

Perceived learning across study groups

Perceived learning outcomes were assessed across four domains: understanding (Q1), heart-sound learning (Q2), lung-landmark learning (Q3), and fun/engagement (Q4). Table 3 presents the descriptive statistics for each domain and the composite perceived-learning scores across the eight study groups.

Overall, participants reported favourable learning experiences, with a composite mean score of 4.40 (SD = 0.80) on the five-point Likert scale. At the domain level, perceived understanding received the highest mean rating (M = 4.60, SD = 0.80), followed by fun/engagement (M = 4.44, SD = 1.03), lung-landmark learning (M = 4.33, SD = 1.01), and heart-sound learning (M = 4.23, SD = 1.01).

Descriptive comparisons revealed variation across the study groups (Table 3). Group C recorded the highest mean ratings for understanding (M = 4.81, SD = 0.40) and heart-sound learning (M = 4.63, SD = 0.50), whereas Group E recorded the highest ratings for lung-landmark learning (M = 4.75, SD = 0.45) and fun/engagement (M = 4.88, SD = 0.34).

The distribution of composite perceived-learning scores across the study groups is illustrated in Fig. 2. Group C recorded the highest composite mean score (M = 4.70, SD = 0.26), followed by Groups E (M = 4.67, SD = 0.24) and D (M = 4.63, SD = 0.50). Group H recorded the lowest composite mean score (M = 4.05, SD = 1.11).

Figure 3 illustrates the group-level distribution of scores across the four perceived-learning domains, showing the differences in participants' ratings of understanding, heart-sound learning, lung-landmark learning, and fun/engagement.

Table 3
Perceived learning scores by study group, mean (SD) on a five-point Likert scale.

Group	Q1 Understanding	Q2 Heart sounds	Q3 Lung landmarks	Q4 Fun/ engagement	Composite
A (control)	4.57 (0.93)	4.10 (1.18)	4.05 (1.43)	4.10 (1.41)	4.20 (1.04)
B	4.31 (1.01)	4.13 (1.03)	4.13 (0.96)	4.63 (1.03)	4.30 (0.97)
C	4.81 (0.40)	4.63 (0.50)	4.69 (0.48)	4.69 (0.48)	4.70 (0.26)
D	4.75 (0.45)	4.38 (0.96)	4.69 (0.48)	4.69 (0.60)	4.63 (0.50)
E	4.75 (0.58)	4.31 (0.60)	4.75 (0.45)	4.88 (0.34)	4.67 (0.24)
F	4.62 (0.62)	4.25 (1.00)	4.13 (1.03)	4.56 (0.51)	4.39 (0.54)
G	4.63 (1.03)	4.00 (1.27)	4.44 (1.09)	4.25 (1.07)	4.33 (0.98)
H	4.37 (1.03)	4.13 (1.31)	3.88 (1.26)	3.81 (1.56)	4.05 (1.11)
Total	4.60 (0.80)	4.23 (1.01)	4.33 (1.01)	4.44 (1.03)	4.40 (0.80)

Scores were based on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Values are mean (SD) scores. The composite score was calculated as the mean of Q1–Q4 for each participant. Group A was the control group; Groups B–H were intervention groups (n = 21 for Group A and n = 16 for each intervention group).

Multivariate and univariate analyses of perceived learning

Table 4 presents the summary of multivariate, conventional univariate, and robust ANOVA findings.

Multivariate analysis of variance (MANOVA) was conducted to examine differences in the four perceived-learning domains across the eight study groups. Box's M test was significant [Box's M = 215.298, F(70, 20402.649) = 2.745, $p < 0.001$], indicating heterogeneity of covariance matrices. Consequently, Pillai's trace was selected as the primary multivariate test statistic. The analysis revealed no statistically significant overall difference in perceived-learning outcomes across the study groups [Pillai's trace = 0.252, F(28, 500) = 1.201, $p = 0.222$, partial $\eta^2 = 0.063$].

Levene's test did not indicate significant heterogeneity of variance for understanding (Q1; $p = 0.346$) or heart-sound learning (Q2; $p = 0.433$), but indicated significant heterogeneity for lung-landmark learning (Q3; $p = 0.023$) and fun/engagement (Q4; $p < 0.001$). However, the conventional univariate ANOVA identified a nominally significant difference in fun/engagement across the eight study groups [F(7, 125) = 2.189, $p = 0.039$].

Because the homogeneity-of-variance assumption was violated for Q3 and Q4, Welch's robust ANOVA was performed. The analysis identified nominally significant differences across the study groups in lung-landmark learning [F(7, 52.709) = 2.509, $p = 0.027$] and fun/engagement [F(7, 52.523) = 2.293, $p = 0.041$]. Supplementary Brown–Forsythe tests yielded a nonsignificant result for Q3 [F(7, 91.339) = 2.073, $p = 0.054$] and a nominally significant result for Q4 [F(7, 74.595) = 2.299, $p = 0.035$].

However, after applying a Bonferroni correction for the four perceived-learning domains (adjusted significance threshold = 0.0125), neither Q3 nor Q4 remained statistically significant. Furthermore, Games–Howell post-hoc comparisons, comprising 28 pairwise comparisons per outcome across the eight study groups, identified no statistically significant between-group differences for any of the four perceived-learning domains. For Q1, all pairwise comparisons were non-significant ($p = .608–1.000$); the largest observed difference was between Groups B and C [mean difference = − 0.50, SE = 0.273, 95% CI (− 1.42, 0.42), $p = .60$]. Similarly, for Q2, all pairwise p-values ranged from .592 to 1.000, with the largest difference observed between Groups C and G [mean difference = 0.62, SE = 0.340, 95% CI (− 0.52, 1.77), $p = .604$]. Although Welch's ANOVA yielded nominal p-values of .027 for Q3 and .041 for Q4, these did not remain significant after Bonferroni adjustment (adjusted $p = .108$ and .164, respectively), and no Games–Howell pairwise comparison was significant (Q3: $p = .209–1.000$; Q4: $p = .202–1.000$). For Q3, the largest difference was between Groups E and H [mean difference = 0.87, SE = 0.334, 95% CI (− 0.26, 2.01), $p = .209$]. For Q4, the largest difference was likewise between Groups E (4.88 ± 0.34) and H (3.81 ± 1.55), with a mean difference of 1.06 ($p = .202$). Overall, the post hoc findings indicate that none of the observed pairwise differences reached statistical significance.

Comparison of composite perceived-learning scores across study groups

Composite perceived-learning scores were compared across the eight study groups. Levene's test indicated unequal variances, $F(7, 125) = 3.090$, $p = 0.005$. The conventional one-way ANOVA showed no statistically significant difference in composite scores across the groups, $F(7, 125) = 1.489$, $p = 0.177$, $\eta^2 = 0.077$.

To account for the unequal variances, Welch's robust ANOVA was performed, which also yielded a nonsignificant result, $F(7, 52.027) = 2.141$, $p = 0.055$. Similarly, Games–Howell post hoc comparisons identified no statistically significant pairwise differences between the groups ($p > .05$). Although descriptive variations in composite perceived-learning scores were observed across the study groups (Fig. 2), these differences were not statistically significant.

Table 4
Multivariate, univariate and robust analyses of perceived-learning outcomes across the eight study groups.

Outcome/test	Test statistic	df	*p*-value	Bonferroni-adjusted *p*	Effect size	Levene's *p*
Multivariate analyses						
Pillai's trace	0.252; F = 1.201	28, 500	0.222	—	0.063	—
Wilks' lambda	0.767; F = 1.205	28, 441.299	0.219	—	0.064	—
Univariate analyses						
Q1: Understanding	F = 0.804	7, 125	0.586	1.000	0.043	0.346
Q2: Heart-sound learning	F = 0.616	7, 125	0.742	1.000	0.033	0.433
Q3: Lung-landmark learning (conventional ANOVA)	F = 1.965	7, 125	0.065	—	0.099	0.023
Q3: Welch's ANOVA	F = 2.509	7, 52.709	0.027	0.108	—	—
Q4: Fun/engagement (conventional ANOVA)	F = 2.189	7, 125	0.039	—	0.109	< 0.001
Q4: Welch's ANOVA	F = 2.293	7, 52.523	0.041	0.164	—	—
Composite perceived-learning score						
Conventional one-way ANOVA	F = 1.489	7, 125	0.177	—	0.077	0.005
Welch's ANOVA	F = 2.141	7, 52.027	0.055	—	—	—

Box's M = 215.298, F(70, 20402.649) = 2.745, p < 0.001, indicating heterogeneity of covariance matrices; therefore, Pillai's trace was used as the primary multivariate statistic. Levene's tests were based on the mean. Welch's robust ANOVA was used for Q3, Q4 and the composite score because of unequal variances. A Bonferroni correction was applied across the four individual perceived-learning domains, yielding an adjusted significance threshold of 0.0125. Neither Q3 nor Q4 remained statistically significant after this correction. Games–Howell post-hoc comparisons identified no statistically significant pairwise differences for Q3, Q4, or the composite score (all p > .05). Effect sizes are reported as partial η^2 for the multivariate and individual univariate analyses and η^2 for the composite-score analysis.

Comparison of composite perceived-learning scores by programme of study

An exploratory Mann–Whitney U test was conducted to compare composite perceived-learning scores between Medicine and Dental Surgery students. Medicine students (n = 108) had a mean rank of 67.14, compared with 66.38 for Dental Surgery students (n = 25).

The analysis revealed no statistically significant difference in composite perceived-learning score ranks between the two programmes (U = 1334.50, Z = − 0.091, p = 0.927), as presented in Table 5.

Table 5
Mann–Whitney U test comparing composite perceived-learning scores between Medicine and Dental Surgery students

Variable	Medicine (n = 108)	Dental Surgery (n = 25)	Test statistic
Mean rank	67.14	66.38	
Sum of ranks	7,251.50	1,659.50	
Mann–Whitney U			1,334.500
Wilcoxon W			1,659.500
Z-value			−0.091
p-value (two-tailed)			0.927

N = 133. The grouping variable was programme of study (Medicine or Dental Surgery).

DISCUSSION

This study evaluated a gamified, wearable human-mannequin approach for teaching thoracic wall surface anatomy and compared students' learning experiences and immediate formative performance across eight instructional groups. Overall, students reported favourable perceptions of their learning experience, with a composite perceived-learning score of 4.40 out of 5. However, neither the multivariate analysis nor the composite analysis demonstrated statistically significant differences among the

instructional groups. Although some item-level variation was observed, post-hoc analyses did not identify reliable pairwise differences. Formative quiz performance was also high across groups, with several items approaching or reaching ceiling levels. Taken together, these findings indicate that the wearable-mannequin approach was well accepted and educationally feasible, but the present data do not demonstrate that it was superior to conventional plastic-model instruction in either perceived learning or objectively measured immediate knowledge.

The generally favourable perceptions are consistent with the broader literature on active learning in medical education. Active-learning strategies are most likely to promote engagement when students participate in authentic tasks, interact with peers, receive timely feedback, and are required to make observable contributions to the learning process [19]. These features were incorporated into the present intervention through role rotation, landmark palpation, surface tracing, peer interaction, colour-coded formative responses, and immediate feedback. Unlike a conventional demonstration in which students may remain relatively passive, the mannequin activity required learners to identify structures actively and translate anatomical knowledge to the surface of a living body. This type of participation is consistent with evidence that active learning can engage behavioural, cognitive, and emotional dimensions of the learning process [19].

Gamification may also help explain the favourable student experience. Systematic evidence from health professions education suggests that gamification can increase participation and motivation, although learning effects vary substantially according to the game elements used, learning objectives, learner characteristics, and assessment methods [9]. Aster et al. [20] similarly showed that challenge, feedback, goals, and other game-design elements are commonly incorporated into serious games in medical and healthcare education, but their effectiveness depends on their pedagogical integration rather than on the mere presence of a game-like activity. In the current study, the colour-coded response system, defined student roles, immediate feedback, and collaborative activity provided elements of challenge, interaction, and participation without separating the game component from the anatomical learning objective.

Recent evidence also supports the potential usefulness of gamified formative assessment in foundational medical sciences. For example, Aldalou et al. [21] reported high learner participation in a gamified formative question bank used among first-year medical students and demonstrated relationships between patterns of engagement and academic performance. Nevertheless, favourable engagement should not be interpreted automatically as evidence of superior learning. A recent meta-analysis of gamification for clinical reasoning found positive pooled effects but also substantial heterogeneity, potential publication bias, and overall low certainty of evidence, indicating that effects may differ considerably between educational settings and intervention designs [22]. This is important in interpreting the present study: the mannequin activity was engaging and positively received, but the statistical findings do not establish a corresponding advantage in knowledge acquisition.

Self-determination theory provides one possible explanation for the positive learner experience. According to Ryan and Deci [23], intrinsic motivation is supported when learning environments facilitate

autonomy, competence, and relatedness. The role-based structure of the present intervention may have supported autonomy by requiring students to perform defined tasks, competence through immediate application and feedback, and relatedness through peer interaction. Consistent with this interpretation, a systematic review and meta-analysis found that gamification can improve intrinsic motivation and perceptions of autonomy and relatedness, although its effect on perceived competence was comparatively small [24]. These mechanisms were not measured directly in the present study and should therefore be regarded as plausible explanations rather than demonstrated mediators.

Immediate feedback was another potentially valuable feature of the instructional design. Students responded to formative questions using colour-coded cards and received feedback immediately after each response. Timely feedback allows learners to compare their reasoning with the correct response while the underlying thought process remains recent. In a randomised study, Luo P. et al.[25] found that formative objective structured clinical examination with immediate feedback improved medical students' self-confidence and selected measures of clinical competence. Although the present colour-card exercise was much simpler than an OSCE, both approaches illustrate the educational value of embedding feedback directly within formative activity rather than separating assessment from learning.

The wearable mannequin may be particularly suitable for surface anatomy because it connects anatomical knowledge with the living body. Living-anatomy teaching has long been proposed as a means of helping students translate structures learned from diagrams, cadavers, and models into clinically identifiable landmarks. Collett et al. [8] described the value of living models in teaching surface anatomy, particularly for relating anatomical structures to features that can be observed or palpated during clinical examination. More recently, Bakar et al. [5] developed a contextualised work-based surface-anatomy teaching model intended to promote clinically relevant learning competencies. These approaches support the principle underlying the present intervention: thoracic landmarks are not merely structures to be recognised on a plastic model but locations that students must eventually identify accurately on patients.

The mannequin activity may therefore have particular relevance to the transition from preclinical anatomy to cardiovascular and respiratory examination. Students palpated and marked structures such as the sternal angle, intercostal spaces, and midclavicular line before identifying cardiac and pulmonary auscultation areas. Simulation literature provides support for combining anatomical location with examination practice. Patrizio et al. [26], in a systematic review of cardiac-auscultation simulation, found that most included studies reported improvements in outcomes such as diagnostic accuracy, knowledge, competence, or learner satisfaction, although considerable heterogeneity existed across interventions and assessments. More broadly, technology-enhanced simulation has demonstrated beneficial effects across health professions education compared with no intervention and, in some contexts, with conventional instruction [12]. Simulation accompanied by deliberate practice has also demonstrated stronger skill-acquisition outcomes than traditional clinical education in meta-analytic evidence [27]. The current findings therefore fit within a broader literature suggesting that simulation can be educationally valuable while not guaranteeing superiority for every outcome or instructional context.

The intervention also has a plausible experiential-learning basis. Experiential learning emphasises that clinically meaningful learning can emerge from participation in authentic or contextually relevant experiences, particularly when experience is accompanied by feedback, reflection, and opportunities to apply knowledge [14]. In this study, students alternated between acting as the mannequin, identifying landmarks, tracing structures, and practising examination-related tasks. Such role rotation may have enabled learners to encounter the same anatomical concepts from different perspectives. However, experiential processes such as reflection, cognitive engagement, and transfer were not measured directly; consequently, the present study cannot determine which component of the activity was responsible for the favourable learner perceptions.

The item-level findings further emphasise the need for cautious interpretation. For fun and engagement (Q4), Group E recorded the highest mean score (M = 4.88, SD = 0.34), whereas Group H recorded the lowest (M = 3.81, SD = 1.56). Although the conventional univariate test produced p = .039, the assumption of equal variances was not satisfied and no Games–Howell pairwise comparison reached statistical significance. The data therefore indicate heterogeneity in reported engagement across groups, but they do not identify a particular group or instructional condition as superior. Moreover, multiple item-level tests increase the probability of chance findings, strengthening the rationale for interpreting the adjusted analyses rather than isolated uncorrected p values.

A similarly cautious interpretation is appropriate for lung-landmark learning (Q3). The omnibus result did not reach conventional statistical significance (p = .065), although its partial η^2 of .099 indicated one of the larger observed item-level effect sizes. A non-significant result should not be described as demonstrating an intervention effect. Nevertheless, because spatial localisation of auscultation landmarks is closely aligned with the activities performed using the wearable model, this domain may warrant specific evaluation in a larger, appropriately powered trial. Such a study should define landmark-identification performance prospectively as an objective outcome rather than relying principally on students' perceptions of their learning.

The current findings can also be considered alongside recent anatomy-education experience at the University of Rwanda. Archibong et al.[15] reported favourable student perceptions of the Anatomage Table at the same university, including perceived improvements in visualisation, understanding of anatomical relationships, and enjoyment of learning. The Anatomage Table and wearable mannequin are fundamentally different instructional technologies, and findings from one cannot be assumed to apply to the other. Nevertheless, both illustrate increasing interest in supplementing conventional anatomy instruction with approaches that make three-dimensional and clinically relevant relationships more tangible. The wearable mannequin may offer an additional modality because it incorporates palpation and surface localisation on a living person rather than visualisation alone.

Objective quiz performance was high. At the individual level, accuracy was approximately 97% for the sternal angle item, 98% for the right recurrent laryngeal nerve item, and 99% for the middle-lobe auscultation item, while the aortic auscultation item produced a lower, though still high, accuracy of 86%.

In the separately scored group-level assessment, three of four questions produced maximum scores in every group. Only the question concerning the aortic auscultation site differentiated the groups descriptively, with Groups B–F scoring 100% and Groups A, G, and H scoring 75%. Because the group-level scores were aggregate scores rather than participant-level observations, inferential comparisons cannot appropriately determine whether this pattern reflects a true intervention effect.

The concentration of scores near the maximum suggests an important ceiling effect. Ceiling effects reduce an instrument's ability to discriminate between learners or detect educational differences because participants with different underlying levels of knowledge may obtain the same maximum or near-maximum score [28]. This problem is particularly relevant here because only four quiz items were available and three generated perfect group-level scores. Students also had access to standardised preparatory materials seven days before the teaching session. This preparation was educationally appropriate but may have elevated baseline knowledge in both the control and intervention conditions, further reducing the ability of a short immediate post-session quiz to distinguish between instructional approaches.

The findings also illustrate why perceived learning and demonstrated learning should be considered separately. High satisfaction or perceived understanding cannot by itself establish greater knowledge acquisition. Deslauriers et al. [29] experimentally demonstrated that students' perception of how much they learned can diverge from objectively measured learning, underscoring the limitations of relying on subjective evaluations as proxies for educational effectiveness. In the present study, Groups C, D, and E combined relatively high perception scores with high quiz performance, whereas Groups B and F achieved maximum group-level quiz scores despite somewhat lower composite perception means of 4.30 and 4.39, respectively. This imperfect alignment reinforces the value of evaluating affective outcomes, perceived understanding, factual knowledge, clinical performance, and longer-term retention as distinct constructs.

The composite perceived-learning analysis should likewise be interpreted conservatively. Although Groups C, E, and D had the highest descriptive means, the conventional ANOVA was not statistically significant. Because the assumption of homogeneity of variance was not satisfied, Welch's test was an appropriate robustness check; this analysis also remained above the conventional significance threshold ($p = .055$). Furthermore, no Games–Howell pairwise comparison was statistically significant. Consequently, the data do not provide sufficient evidence that overall perceived learning differed reliably among the eight groups. Conversely, failure to demonstrate a difference does not establish equivalence or non-inferiority between the mannequin and conventional approaches. Such conclusions require an adequately powered equivalence or non-inferiority design with a prespecified margin.

An additional analytical consideration is that Groups B–H all received the wearable-mannequin intervention, whereas Group A alone served as the conventional-model control. Accordingly, comparisons among all eight groups primarily describe variability between instructional groups and do not represent the most direct statistical test of the study's principal intervention question. If participant-

level outcome data are available, an analysis comparing the pooled intervention condition (Groups B–H) with the control condition (Group A), while appropriately considering any clustering associated with instructional groups, would provide a more direct estimate of the intervention effect. The present eight-group analyses remain useful descriptively but should not be interpreted as equivalent to a properly specified intervention-versus-control analysis.

The demographic profile should also be considered when assessing generalisability. Approximately 63.9% of participants were male, and 81.2% were medical students, leaving a substantially smaller proportion of dental surgery students. No difference in composite perceived learning was detected between medical and dental students, but the unequal subgroup sizes limit the precision of this comparison. Furthermore, because allocation was not reported as stratified by sex, programme, or prior academic performance, balance in these characteristics across the eight groups cannot be assumed. Future studies should consider stratified randomisation or covariate-adjusted analyses where appropriate.

From an educational perspective, the findings support viewing the wearable mannequin as a supplementary rather than replacement approach. Its main pedagogical distinction is the combination of living surface anatomy, peer-assisted activity, role rotation, immediate feedback, and simulated clinical examination within a single teaching session. Active-learning strategies can be particularly useful when tasks are relevant, participatory, collaborative, and accompanied by appropriate feedback Luo et al. (2026). However, the present study did not assess costs, implementation burden, instructor training requirements, durability of materials, or cost-effectiveness. Claims that the mannequin is necessarily cheaper or more scalable than alternative methods should therefore be tested rather than assumed.

Importantly, the wearable model should not be positioned as a substitute for established anatomy-teaching resources such as cadaveric material, prosections, imaging, conventional models, or digital three-dimensional resources. Rather, it may complement these modalities by addressing a specific educational need: translating structural anatomy into palpable and clinically meaningful surface landmarks. The findings from this study provide preliminary support for feasibility and acceptability but not evidence of statistical superiority.

LIMITATIONS AND FUTURE DIRECTIONS

Several limitations affect interpretation of the study. First, the post-test-only design did not include baseline measurement of anatomical knowledge, landmark-identification ability, engagement, or perceptions. Therefore, pre-existing differences between groups cannot be excluded, even though students were allocated to groups on the day of teaching.

Second, the source population consisted of 170 eligible students, but the 133 participants were recruited on the basis of availability and consent. Thus, random allocation to instructional groups occurred after recruitment from a convenience-based participant pool. This distinction is important because random

group allocation does not eliminate potential selection differences between participants and non-participants.

Third, the intervention and control structure was unbalanced, with one control group of 21 students and seven intervention groups of 16 students each. Analysis of eight groups separately may therefore be less informative for the primary study question than a prespecified comparison of the intervention and control conditions. Future trials should define the unit of randomisation, intervention contrast, and primary statistical model before data collection. Where teaching occurs in small groups, potential clustering should be incorporated into sample-size calculations and analyses.

Fourth, group sizes were relatively small, reducing precision, particularly for pairwise comparisons. Violations of the homogeneity-of-variance assumption were also observed for several outcomes, although robust procedures such as Welch's test and Games–Howell comparisons appropriately reduced reliance on conventional equal-variance assumptions.

Fifth, the objective assessment was limited. The formative quiz contained only four items, several of which produced near-perfect or perfect scores. This restricted range created a substantial risk of ceiling effects and limited the capacity to distinguish levels of learning. Moreover, group-level quiz scores were available only as aggregates and were not derived in a way that permitted participant-level inferential analysis. Future studies should retain individual-level responses and use a larger, blueprint-driven assessment containing items of varying difficulty. Where possible, reliability and validity evidence should be established for both the knowledge test and perception instrument.

Sixth, the outcomes were assessed immediately after the intervention. Immediate performance cannot determine whether learning is retained, whether students can accurately identify landmarks several weeks later, or whether they can transfer anatomical knowledge to examination of real patients. Follow-up assessments are therefore necessary.

Seventh, perceived-learning outcomes were self-reported and may have been influenced by novelty, enthusiasm for an unfamiliar instructional method, peer interaction, or social desirability. The distinction between perceived and objectively demonstrated learning should therefore remain explicit.

Finally, the study was conducted at a single institution, limiting external generalisability. The effectiveness and feasibility of the wearable approach may differ according to class size, teaching resources, curriculum organisation, faculty experience, cultural context, and prior exposure to active-learning methods.

Future studies should use a clearly specified randomised controlled design with baseline and post-intervention measurements, participant-level objective outcomes, and a priori designation of a primary endpoint. Objective assessment should extend beyond factual multiple-choice questions to include direct identification of thoracic landmarks, structured palpation tasks, recognition of cardiac and respiratory auscultation sites, and, where feasible, OSCE-style clinical examination. Delayed

reassessment would establish whether learning is retained. Larger multicentre studies would improve generalisability and permit prespecified subgroup analyses.

Future research should also separate the components of the intervention. The wearable apron, peer-assisted learning, role rotation, gamification, formative questioning, immediate feedback, and simulated auscultation were delivered together, so the present study cannot establish which element contributed most to engagement or learning. Factorial or sequential studies may help determine whether the wearable model itself provides additional benefit beyond active-learning and peer-teaching strategies. Implementation outcomes (including usability, faculty workload, student acceptability, material costs, fidelity, and sustainability) should also be assessed before widespread curricular adoption.

Overall, the findings suggest that gamified wearable-mannequin instruction is a feasible and well-received method for contextualising thoracic surface anatomy and linking anatomical landmarks with clinical examination. However, the absence of statistically reliable between-group differences, the ceiling effect in objective assessment, and the limitations of the post-test-only design mean that superiority over conventional plastic-model instruction cannot presently be claimed. The appropriate interpretation is therefore that the approach is a promising supplementary teaching strategy whose effects on objective learning, retention, and clinical performance require confirmation in a more rigorous trial.

CONCLUSION

Students in all eight instructional groups reported favourable perceptions of learning thoracic surface anatomy and cardiac and pulmonary auscultation, and performance on the gamified formative quiz was high. However, the multivariate and composite analyses did not demonstrate statistically significant differences between groups; the omnibus effect on fun/engagement was not localised to any pair of groups; and the descriptive difference on a single quiz item cannot be attributed to the intervention. The findings therefore support the feasibility and acceptability of interactive, mannequin-assisted and gamified teaching, but they do not establish its statistical superiority or equivalence to conventional plastic-model teaching.

Given its low cost and scalability, the wearable thoracic mannequin apron is a reasonable adjunct to conventional anatomy teaching, particularly in resource-constrained settings. Larger, more rigorously controlled studies with baseline assessment, participant-level objective scores, more discriminating assessment items and longer-term measures of clinical skill and knowledge retention are required to determine which components of this approach produce durable improvements in learning.

Declarations

Ethics approval:

Ethical approval (No. 509/CMHS IRB/2024) was obtained from the Institutional Review Board (IRB) of the College of Medicine and Health Sciences, University of Rwanda, before the study began. All

participants were treated in accordance with the ethical principles of the World Medical Association Declaration of Helsinki (2024).

Informed Consent

Participants were informed of the aim of the research and gave written informed consent voluntarily after reading the consent form.

Consent for publication:

Not Applicable

Competing interests:

All authors declare no conflict of interest.

Clinical Trial Number

Not Applicable

Funding statement:

None

Author Contribution

VBA, PMK and JG, conceptualized the study. VBA, and PMK designed the study. VBA, FA, HA, SO, UI, AM, and CW collected data. VBA, EME, UI, FA, and HA conducted statistical analysis. VBA, FA, HA, SO, UI, AM, CW, PMK, DCI, and EME drafted the initial manuscript. All authors approved the final version for publication

Acknowledgement

The authors would like to thank the students who consented to be part of this study.

Data Availability

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

References

1. Estai M, Bunt S. Best teaching practices in anatomy education: A critical review. Annals Anat - Anatomischer Anzeiger. 2016;208:151–7. https://doi.org/10.1016/J.AANAT.2016.02.010.
2. Syed Abd Halim SA, Yusoff MSB, Yaman MN, Razali SA, Tengku Muda TFM, Ramli RR, et al. Clinical students' reflections on the preclinical anatomy learning experience. J Taibah Univ Med Sci. 2023;18:757–70. https://doi.org/10.1016/J.JTUMED.2022.12.007.
3. Freeman S, Eddy SL, McDonough M, Smith MK, Okoroafor N, Jordt H, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A. 2014;111:8410–5. https://doi.org/10.1073/PNAS.1319030111.
4. Prince M. Does active learning work? A review of the research. J Eng Educ. 2004;93:223–31. https://doi.org/10.1002/J.2168-9830.2004.TB00809.
5. Bakar YIA, Yusoff MSB, Kasim F, Sulong HAM, Hadie SNH. Work-based Surface Anatomy Teaching Model: A Contextualized Method for Promoting Learning Competencies in Medical Curricula. Med Sci Educ 2025. 2025;35:4. https://doi.org/10.1007/S40670-025-02385-W.
6. Boon JM, Meiring JH, Richards PA. Clinical anatomy as the basis for clinical examination: Development and evaluation of an Introduction to Clinical Examination in a problem-oriented medical curriculum. Clin Anat. 2002;15:45–50. https://doi.org/10.1002/CA.1091.
7. Zhang Y, Ji Z, Zhou P, Dong L, Chen Y. Clinical anatomy teaching: A promising strategy for anatomic education. Heliyon. 2023;9. https://doi.org/10.1016/j.heliyon.2023.e13891.
8. Collett T, Kirvell D, Nakorn A, McLachlan JC. The role of living models in the teaching of surface anatomy: Some experiences from a UK Medical School. Med Teach. 2009;31. https://doi.org/10.1080/01421590802516731.
9. van Gaalen AEJ, Brouwer J, Schönrock-Adema J, Bouwkamp-Timmer T, Jaarsma ADC, Georgiadis JR. Gamification of health professions education: a systematic review. Adv Health Sci Educ. 2021;26:683–711. https://doi.org/10.1007/S10459-020-10000-3.
10. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online. 2018;23. https://doi.org/10.1080/10872981.2018.1438718.
11. Huang WD, Loid V, Sung JS. Reflecting on gamified learning in medical education: a systematic literature review grounded in the Structure of Observed Learning Outcomes (SOLO) taxonomy 2012–2022. BMC Med Educ 2024. 2024;24:1. https://doi.org/10.1186/S12909-023-04955-1.
12. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-Enhanced Simulation for Health Professions Education: A Systematic Review and Meta-analysis. JAMA. 2011;306:978–88. https://doi.org/10.1001/JAMA.2011.1234.
13. Kumarasamy PM. Human mannequin: Learning anatomy and clinical skills via surface marking of the chest wall, a distinctive non-cadaveric approach (A-0374). In: ACB Journditor. IFAA 2024. 2024.

14. Yardley S, Teunissen PW, Dornan T. Experiential learning: AMEE Guide 63. Med Teach. 2012;34. https://doi.org/10.3109/0142159X.2012.650741.
15. Archibong VB, Olorunnado S, Mohammed A, Okesina A, Ndahimana P, Habumuremyi S, et al. Students' opinions on the use of the anatomage table in anatomy learning at the University of Rwanda. Discover Educ 2025. 2025;4:1. https://doi.org/10.1007/S44217-025-00840-9.
16. Rwanda Statistical Yearbook – 2019 | National Institute of Statistics of Rwanda. https://statistics.gov.rw/statistical-publications/statistical-yearbook/rwanda-statistical-yearbook-2019. Accessed 24 Sep 2026.
17. Krejcie RV, Morgan DW. Determining Sample Size for Research Activities. Educ Psychol Meas. 1970;30:607–10. https://doi.org/10.1177/001316447003000308.
18. Nyamweya NN, Yekka P, Mubutu RD, Kasozi KI, Muhindo J. Staff Absenteeism in Public Health Facilities of Uganda: A Study in Bushenyi District on Contributing Factors. Open J Nurs. 2017;07:1115–30. https://doi.org/10.4236/OJN.2017.710081.
19. Luo X, Alias BS, Adnan NH. Active learning strategies and student engagement in undergraduate medical education: A systematic review (2015–2025). Anat Sci Educ. 2026;19:1157–69. https://doi.org/10.1002/ASE.70231. ;CTYPE:STRING:JOURNAL.
20. Aster A, Laupichler MC, Zimmer S, Raupach T. Game design elements of serious games in the education of medical and healthcare professions: a mixed-methods systematic review of underlying theories and teaching effectiveness. Adv Health Sci Educ 2024. 2024;29:5. https://doi.org/10.1007/S10459-024-10327-1.
21. Aldalou M, Laboy A, Fraser LE, Walker J, Brooks WS, Hubner B. A gamified formative question bank to enhance student learning and engagement in a foundational medical science course. Anat Sci Educ. 2026;19:891–901. https://doi.org/10.1002/ASE.70107.
22. Lee CY, Lee CH, Lai HY, Chen PJ, Chen MM, Yau SY. Do games work? A meta-analytic synthesis of gamified learning for clinical reasoning in medical and allied health education. Med Educ Online. 2026;31. https://doi.org/10.1080/10872981.2026.2614233.
23. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol. 2000;55:68–78. https://doi.org/10.1037/0003-066X.55.1.68.
24. Li L, Hew KF, Du J. Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: a meta-analysis and systematic review. Educational Technol Res Dev 2024. 2024;72:2. https://doi.org/10.1007/S11423-023-10337-7.
25. Luo P, Shen J, Yu T, Zhang X, Zheng B, Yang J. Formative objective structured clinical examination with immediate feedback improves surgical clerks' self-confidence and clinical competence. Med Teach. 2023;45:212–8. https://doi.org/10.1080/0142159X.2022.2126755.
26. Patrizio HA, Phyu R, Kim B, Brolis NV. Utilization of Simulation to Teach Cardiac Auscultation: A Systematic Review. 2023. https://doi.org/10.7759/cureus.41567

27. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does Simulation-Based Medical Education With Deliberate Practice Yield Better Results Than Traditional Clinical Education? A Meta-Analytic Comparative Review of the Evidence. Acad Med. 2011;86:706–11. https://doi.org/10.1097/ACM.0B013E318217E119.
28. Edwards KD, Soland J. How Scoring Approaches Impact Estimates of Growth in the Presence of Survey Item Ceiling Effects. Appl Psychol Meas. 2024;48:147–64. https://doi.org/10.1177/01466216241238749.
29. Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc Natl Acad Sci U S A. 2019;116:19251–7. https://doi.org/10.1073/PNAS.1821936116.

Figures

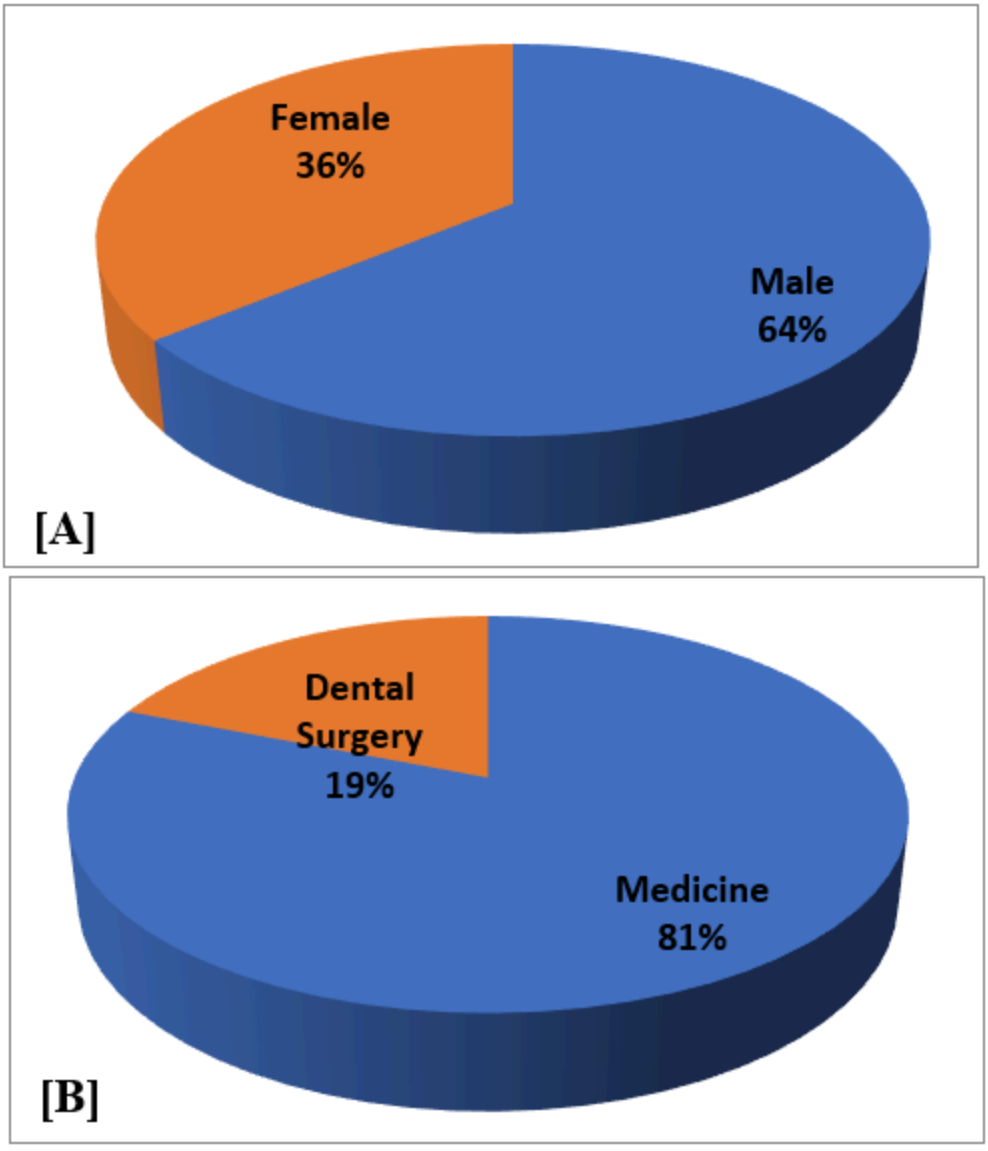


Figure 1

Sociodemographic characteristics of the participants (*N* = 133).

(A) Sex; (B) programme of study.

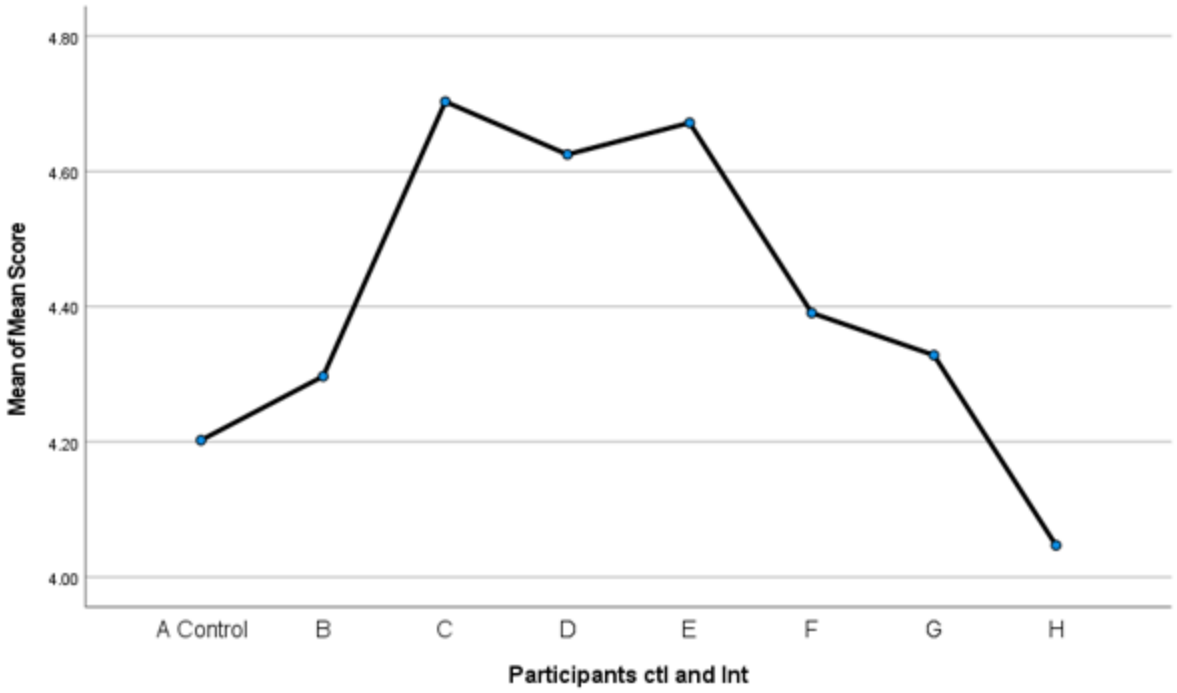


Figure 2

Composite perceived-learning score by study group (mean and 95% confidence interval; *N*= 133). The control group is A, and the intervention groups are B–H.

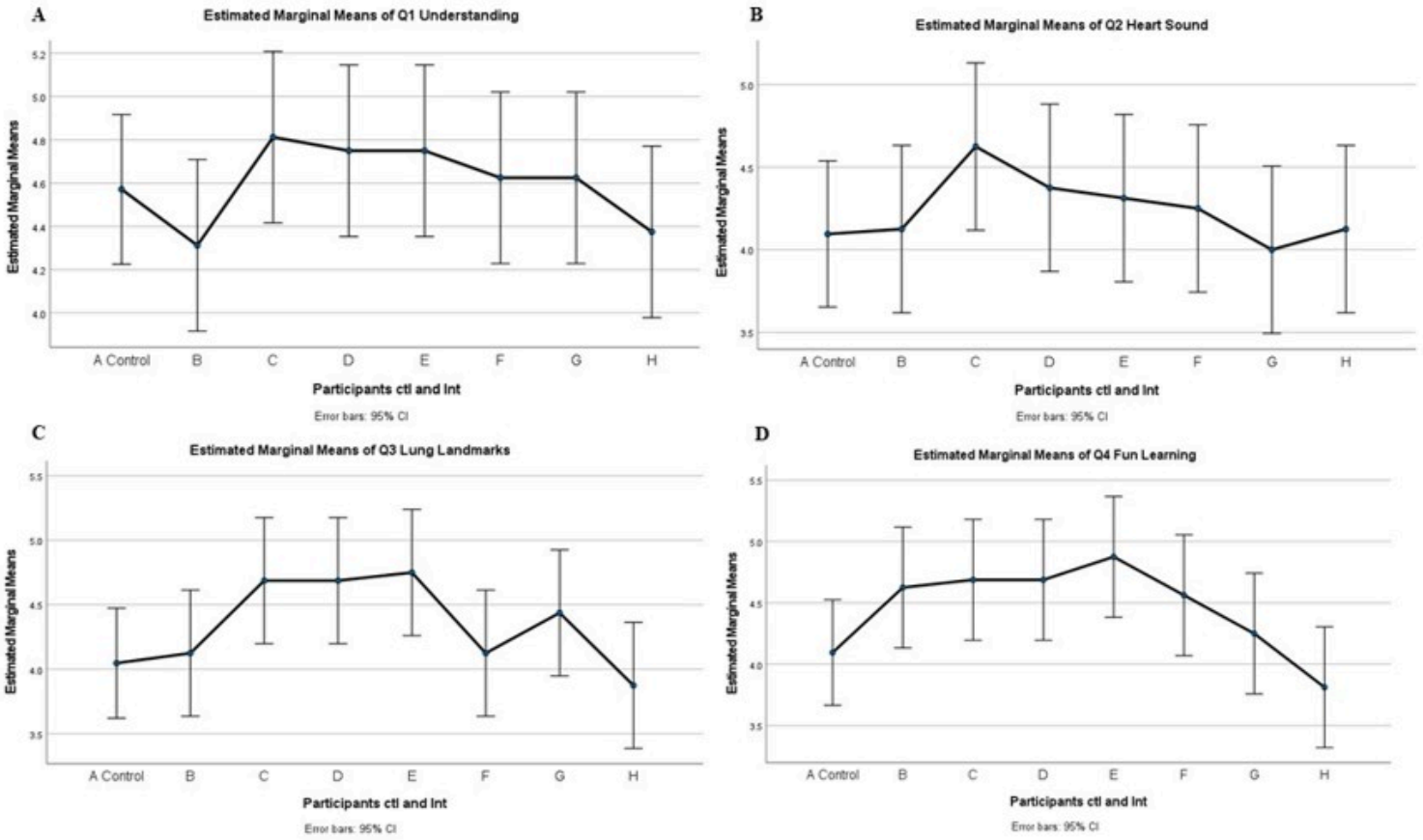


Figure 3

Perceived-learning scores by study group. **A** = Q1, perceived understanding; **B** = Q2, heart-sound learning; **C** = Q3, lung-landmark learning; and **D**= Q4, fun/engagement (mean and 95% confidence interval). Groups: A is the control(ctl), and B–H are the intervention (Int) groups.

Supplementary Files

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

- Gamificationdatacollectiontool.docx