International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-024-03291-z

ORIGINAL ARTICLE

Validity of a virtual reality-based straight coloanal anastomosis simulator

George Westergaard[1] · Alexis Desir[2] · Jacob Barker[3] · Tansel Halic[4] · Shruti Hegde[2] · Amr Al Abbas[2] · Javier Salgado Pogacnik[2] · James W. Fleshman[5] · Ganesh Sankaranarayanan[2] · Suvranu De[6] · Doga Demirel[3]

Received: 13 May 2024 / Accepted: 6 November 2024
© CARS 2024

Abstract

Purpose Current training methods for surgical trainees are inadequate because they are costly, low-fidelity, or have a low skill ceiling. This work aims to expand available virtual reality training options by developing a VR trainer for straight coloanal anastomosis (SCA), one of the Colorectal Objective Structured Assessment of Technical Skills (COSATS) tasks.

Methods We developed a VR-based SCA simulator to evaluate trainees based on their performance. To increase the immersiveness, alongside the VR headset, we used haptics as the primary method of interaction with the simulation. We also implemented objective performance metrics to evaluate trainee performance throughout the simulation.

Results We presented our performance metrics to 27 participants for an Expert Consensus Survey (5-point Likert scale) and created weights for our metrics. The weighted average scores for the 24 task-specific metrics ranged from 3.5 to 5. Additionally, for the general metrics, the scores spanned from 3.3 to 4.6. In the second phase of our study, we conducted a study with 16 participants (novice $n = 9$, expert $n = 7$). Based on the performance, experts outperformed novices by 8.56% when referring to the total score ($p = 0.0041$). Three of the measurable metrics, purse suture ($p = 0.0797$), retracting the anvil ($p = 0.0738$), and inserting the colonoscope ($p = 0.0738$) showed a significant difference between experts and novices. Experts were smoother with their hand motions by 3.67% per second and took 70.77% longer paths to complete the same tasks.

Conclusion We created a high-fidelity coloanal anastomosis VR simulator. The simulator runs in real-time while allowing high immersion with a VR headset, deformable bodies, and a haptic device while providing objective feedback through performance metrics. Experts obtained higher scores throughout the simulation, including the quiz to demonstrate procedural knowledge, the metrics to demonstrate experience in steps/procedure, and control of their basic surgical skills and hand movements.

Keywords Surgical simulator · Coloanal anastomosis · COSATS · Haptics · Virtual reality

Introduction

Colorectal Objective Structured Assessment of Technical Skills (COSATS) is an assessment of technical skills, developed by the American Society of Colon and Rectal Surgeons in cooperation with the American Board of Colon and Rectal Surgery in an attempt to quantify surgical skills for certification [1, 2]. This assessment requires the trainee to complete a circuit of eight different stations, each including a procedure that must be performed. Five of the eight stations are

✉ Doga Demirel
doga@ou.edu

1 Department of Computer Science, Florida Polytechnic University, Lakeland, USA

2 Department of Surgery, University of Texas Southwestern Medical Center, Dallas, USA

3 School of Computer Science, University of Oklahoma, Devon Energy Hall, 110 W. Boyd St, Norman, OK 73019, USA

4 Peachtree Corners, Sunnyvale, USA

5 Department of Surgery, Baylor Scott & White Health, Dallas, USA

6 College of Engineering, Florida A&M University, Florida State University, Tallahassee, USA

Published online: 21 November 2024

open surgical tasks, two are laparoscopic, and one is endoscopic. Simulation and automatic evaluation of these tasks could greatly expedite the training process. This paper covers the implementation of the simulated coloanal anastomosis, one of the five open surgical COSATS tasks: a) linear stapler anastomosis, b) straight coloanal anastomosis (SCA), c) hand-sewn anastomosis, d) ileal pouch-anal anastomosis, and e) rectal prolapse colorectal surgery.

Coloanal anastomosis is primarily used to reconnect the bowel after the resection of low rectal cancers. Anastomotic leak is a significant post-operative concern and is associated with mortality (6% to 22%). [3] This emphasizes the need for proficiency in high-stakes technical procedures in the care of vulnerable patients. Unfamiliarity with a task such as stapling a coloanal anastomosis close to the dentate line can increase the chance of a clinical anastomotic leak [4, 5]. Performing coloanal anastomosis requires proficiency in technique and a good understanding of the circular end-to-end anastomosis (EEA) stapler. Preservation of colonic continuity requires proficiency in placing purse string sutures and careful use of a circular stapler to join the colon and the anus.

Conventional training options are often limited by either cost or fidelity. Cadavers may provide anatomical accuracy but are limited because of tissue changes post-mortem, expense, and limited access. Animal models are cheaper but are less relevant because their anatomy is different [6]. Physical bench models may be inexpensive and adequate in training basic surgical skills [7] but do not teach surgical decision-making. Current virtual reality (VR) training simulations focus on a single aspect of the surgery or sacrifice user immersion for ease of implementation [8–10]. This could impact the quality of training the user receives from interacting with the simulation [11].

We have developed the SCA simulation to provide an immersive experience and increase the quality of the training received. The rationale behind this work relates to increasing the visual fidelity of the virtual world to allow detailed implementation of the integral procedures and boost user immersion and, therefore, improve the quality of training received. To the best of our knowledge, there is not a VR trainer for the open SCA procedure.

The contribution of this work includes: a) a hierarchical task analysis (HTA), b) developing performance metrics, and c) validation of an immersive real-time VR-based SCA simulator with automated objective performance metrics that can differentiate between expert and novice surgeons. User input and haptic feedback are facilitated through haptic devices. A comparative user study was conducted to validate both the simulator and the objective performance metrics.

As simulation technology progresses in accuracy and usability, there has been an increase in applications in the medical field. While tools exist or are being developed to simulate aspects of surgery [8–10], they compromise on user immersion and the provision of a cohesive experience for implementation. Sararit et al. [8] created a VR simulation to train emergency management in dental surgery. One of the main goals was to develop the simulation while minimizing cost to increase accessibility, but this ended up sacrificing immersion and the quality of the experience for the user. Van Nguyen et al. [9] designed and evaluated a VR trainer for needle insertion tasks commonly found in surgery. While their goal was to replicate real-life training methods in a cheaper, virtual environment and measure their efficacy in replicating the training received from such methods, other pressing points were presented. Notably, during the experimental studies where users were tasked with inserting a needle under varying conditions, users performed better when the VR setup had additional features that increased the realism of the simulation. This is partly due to the additional reference points users could use to more effectively navigate the virtual world when the more realistic features were utilized.

In Wu et al. [10], researchers developed a minimally invasive surgery (MIS) system for use in VR. To supplement this VR simulator system, researchers also developed a haptic force and friction device that would provide feedback to the user if the user were to approach a vessel wall with the simulated device during surgery. Salleh et al. [12] developed a similar VR simulator incorporating a haptic input device with the goal of MIS training. But the simulation is graphically unrefined; like the previous system, no effectiveness of the device is measured. Chemlal et al. [13] employed a similar haptic device in developing a non-VR surgical simulator for MIS and reported promising results in increasing the realism of the simulation. Kannangara et al. [14] showed value in including haptic devices in the laparoscopic surgery simulation, as users utilizing a haptic input device could correctly identify organs based on haptic feedback. Alvarez-Lopez et al. [15] also received promising results on the opposite end of the realism spectrum. Researchers set out to create a low-cost VR MIS simulator as a training supplement, and users reported that they were confident that it could be used to teach laparoscopic surgery.

Demirel et al. [16] developed a virtual airway skills trainer utilizing VR and haptics. The authors created two simulators in the work: endotracheal intubation and cricothyroidotomy [17]. The authors utilized the haptics to find landmarks and create incisions in cricothyroidotomy [18] and move the patient's neck to find the optimal angle for intubation for endotracheal intubation.

Both Qian et al. [19] and Shi et al. [20] designed a solution for the problem of cutting soft bodies during runtime. While this is a pressing issue in the field of simulation, the proposed solutions would have improved accuracy if they were to be implemented in medical simulations. In Li et al. [21], researchers implemented a regulated training system based on VR technology to train users on performing invasive knee surgery. While they did create a virtual simulator that trained users on the surgery, the paper focused more on the administration of the training system than the implementation of a high-quality, immersive simulation.

Gao et al. [22] developed a particle simulator to simulate blood flow in training simulations involving blood vessels, catheters, and guide wires. The study states it is developed to be used in real-time simulations, but the calculations for the model are entirely done on the CPU. This may pose a problem when attempting to implement it in a surgery simulation, as the processor may be overwhelmed by the strain of simulating both the particle physics and the rest of the virtual scene.

Authors of Vaughan et al. [11] found evidence suggesting that a user's ability to perform in VR surgical simulators translates to surgery. They also found strong evidence suggesting that VR surgical simulators could evaluate a user's skill level. Both of these observations were made for arthroscopy simulation. Wijewickrema et al. [23] corroborated some of these findings, showing effectiveness in using VR training for cochlear implant surgery. Sommer et al. [24] also corroborated these findings but for laparoscopic surgical skills.

Methods

Virtual reality simulation framework

The VR SCA simulator is a collaborative effort between three main components, the I/O devices, the VR simulation, and the metrics, which can be seen in Fig. 1.

To interact with the virtual scene, amplify immersion in the simulation, and increase the skill ceiling users can reach [25], a VR headset and two 3D Systems Touch haptic devices capable of providing force feedback are used as the primary interaction methods with the VR simulator. In the simulation, manipulation of complex objects, such as a suture needle with a line, DeBakey, Hemostat, Allis Clamps, Circular Stapler, Needle Holder, and Insufflation Device, is carried out using the haptic devices.

To create the simulation, we used the unity development platform. To realistically represent deformable objects such as a colon or suture thread, we employed the use of eXtended Position-Based Dynamics (XPBD) [26] to represent these objects in the simulation. The pseudo-decoupling of iteration count from object stiffness offered by XPBD allows for more granular control over the behavior of deformable objects in the simulation while maintaining real-time performance, which was utilized to make the deformable objects behave in a more realistic manner. The positions used to represent the suture line were subject to constraints that made the suture line maintain its rest shape and resist torsion, while the positions used to represent the colon were subjected to volume-maintaining constraints to model it as a soft body.

The VR simulation houses every 3D object the user sees and interacts with. An operation area is provided with preselected tools and the organs (colon and rectal stump) for the procedure. This area also includes instructions to assist with the controls and steps of the operation. The setup in the operation room was finalized after consulting with expert colorectal surgeons.

The final component of the system architecture is the performance metrics [27–30]. Our performance metrics [31] have been derived by creating a hierarchical task tree and talking to expert surgeons for fine-tuning. Performance metrics have been incorporated into the simulator to provide objective feedback at the end of the simulation.

Hierarchical task analysis and the simulation

Within this study, we created a hierarchical task analysis (HTA) tree. The HTA breaks down the complex nature of the procedure into smaller components, which makes it easier to identify and analyze the procedural steps and goals, relationships between the sub-steps, and determine required and optional steps [17, 27, 32–34]. By breaking down the procedure into manageable steps, it becomes easier to design training modules that gradually introduce and reinforce the skills needed to perform each sub-task in the simulation. The HTA created can be seen in Fig. 2.

As seen in the HTA, the SCA procedure has five significant tasks, which were implemented in the simulation. **Position the Colon**: Following the completion of the transection using a straight stapler, the distal end of the colon is opened, and Allis Clamps are employed to maintain the colon in an open position, which can be seen in Fig. 3. **Purse Suture and Anvil Insertion:** A significant component of performing an SCA is carrying out a purse string suture [35] around the opening of the colon. Absorbable monofilament sutures are utilized to create consistent bites through the entire circumference of the seromuscular layer. These sutures are approximately 2 mm apart and equidistant. In the suturing step, the user controls a needle driver holding a threaded needle in one hand and a DeBakey forceps in the other. The user must drive the needle through the colon with the needle driver and then grab it and pull with the DeBakey forceps to make a suture point. Once that is completed, the anvil is inserted into the bowel lumen, and the purse suture is tightened and secured around the anvil base as seen in Fig. 4.

After fully closing the suture line around the anvil, the circular stapler is introduced. **Circle Stapler Positioning**: Gently introducing the stapler through the anal canal, the

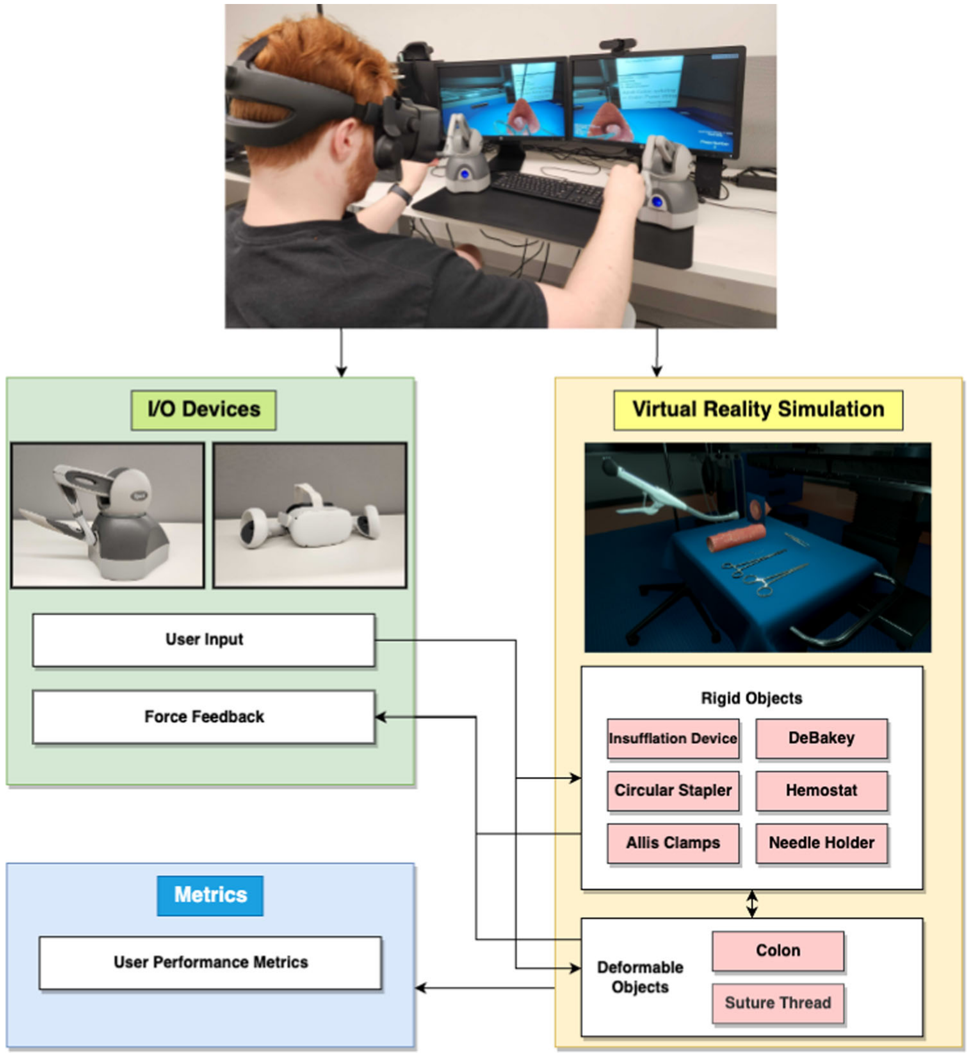

Fig. 1 Overall system architecture

transverse staple line is flattened. The central trocar/spike is deployed, piercing the anal stump until it reaches the anvil in the distal colon. The central trocar/spike and the anvil are carefully connected, and the trocar/spike is retracted to bring the distal colon down to the anus, aligning the tissues. **Staple Firing**: With the circular stapler safety turned off, the stapler is fired to complete the anastomosis. The safety is then turned back on, and the trocar is extended to release the tissue fully. The circular stapler carefully and slowly rotated a small amount and then gently removed through the anus. Circular stapler positioning and staple firing tasks in the simulator are shown in Fig. 5. **Post Firing and Evaluation**: The anastomosis is visually and manually examined for proper alignment, the absence of tension, and perforation or tissue damage. A leak test, derived from the correctness of the anastomosis, is conducted to confirm the absence of anastomotic leaks as seen in Fig. 6.

Expert consensus survey

We carried out an Expert Consensus Survey (IRB # STU-2021–0202) involving 27 colorectal surgeons, with 26 possessing over 5 years of experience and one with 2–5 years of expertise. Participants were asked to rate the clinical significance of each metric item presented in Table 1 on a 5-point Likert Scale, from 1 (not important) and 5 (very important).

User study design and data collection for the simulator

We conducted an institutional review board-approved study (IRB # STU-2021–0202), at the University of Texas Southwestern and Baylor Scott & White Health. In our study, we had a total of 16 participants. After the participants provided informed consent, they were divided into two groups based

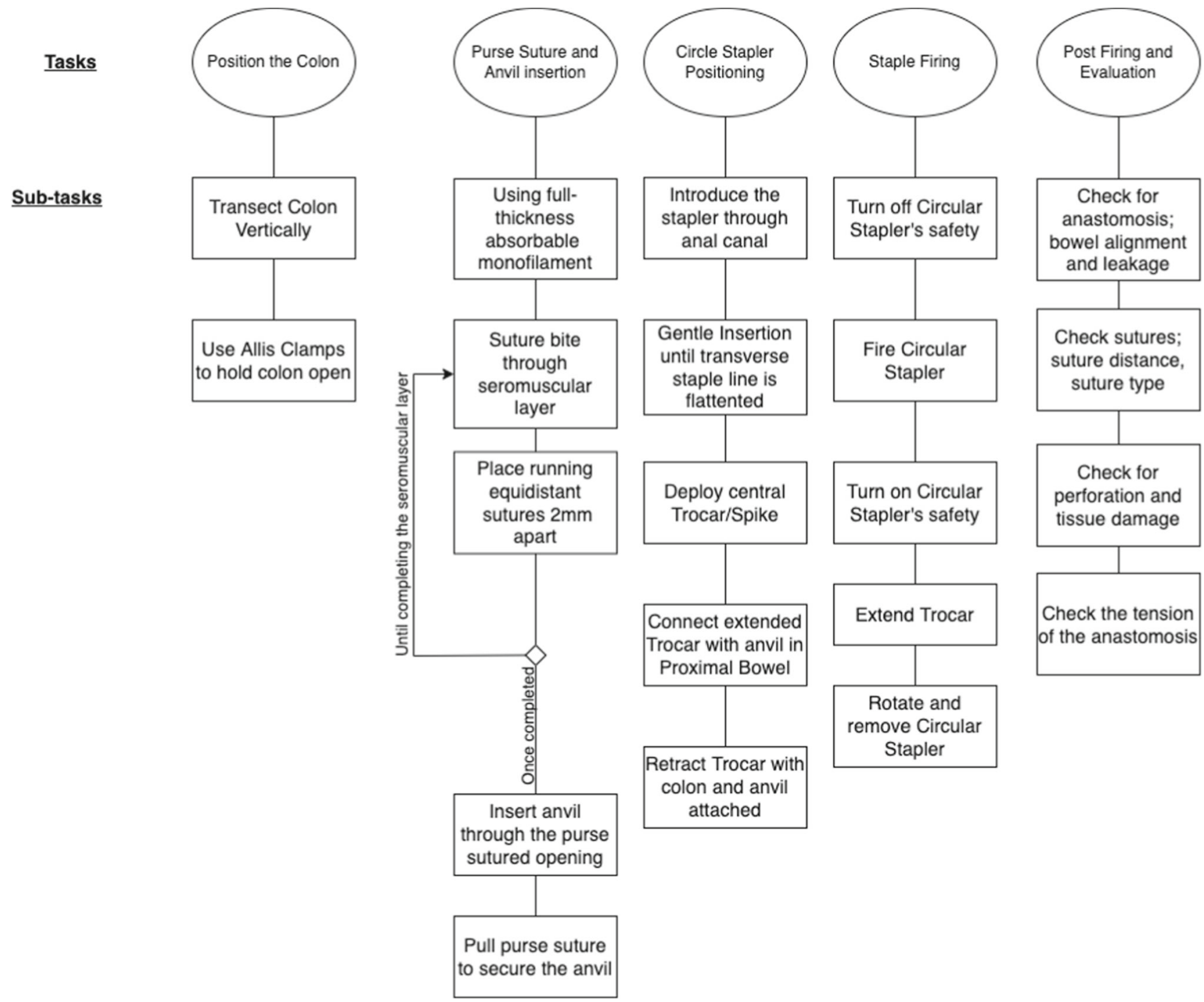

Fig. 2 Hierarchical task analysis tree for the straight coloanal anastomosis procedure

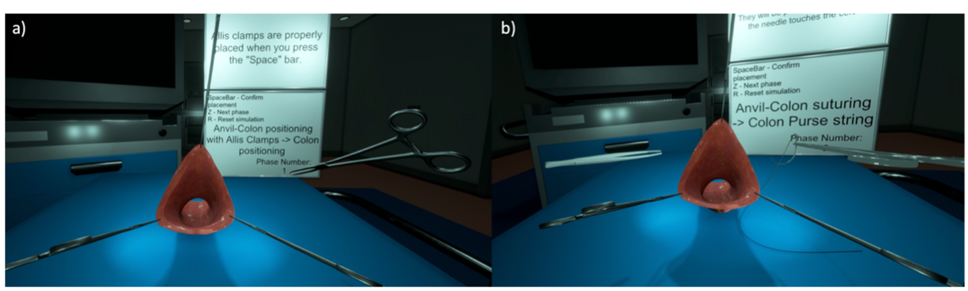

Fig. 3 **a** Positioning of the deformable mesh for purse suture and **b** Start of the purse string suture task

on their experience: experts (n = 7) and novices (n = 9). The novices consisted of PGY 3–5, while experts were either attendings or fellows. Of the novices, five were PGY3, two were PGY4, and two were PGY5. Of the experts, five were attendings, while two were fellows.

For this study, we employed pre- and post-questionnaires, performance metrics (as seen in Table 2), and the data collected from the simulator. In the pre-questionnaire, participants were asked their anonymous demographic information (age, gender) and their experience level (Attending or PGY and year). After filling out the pre-questionnaire, they were given a short orientation about the simulator and the user interface, after which they proceeded to perform the task once on the simulator. In the post-questionnaire, participants

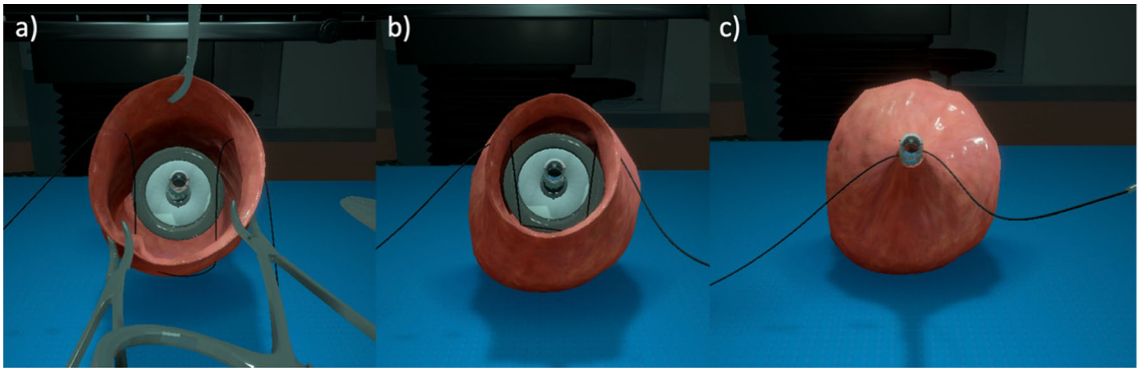

Fig. 4 **a** Anvil placement, **b** Purse string suture pulling, and **c** Fully closed suture line around the anvil

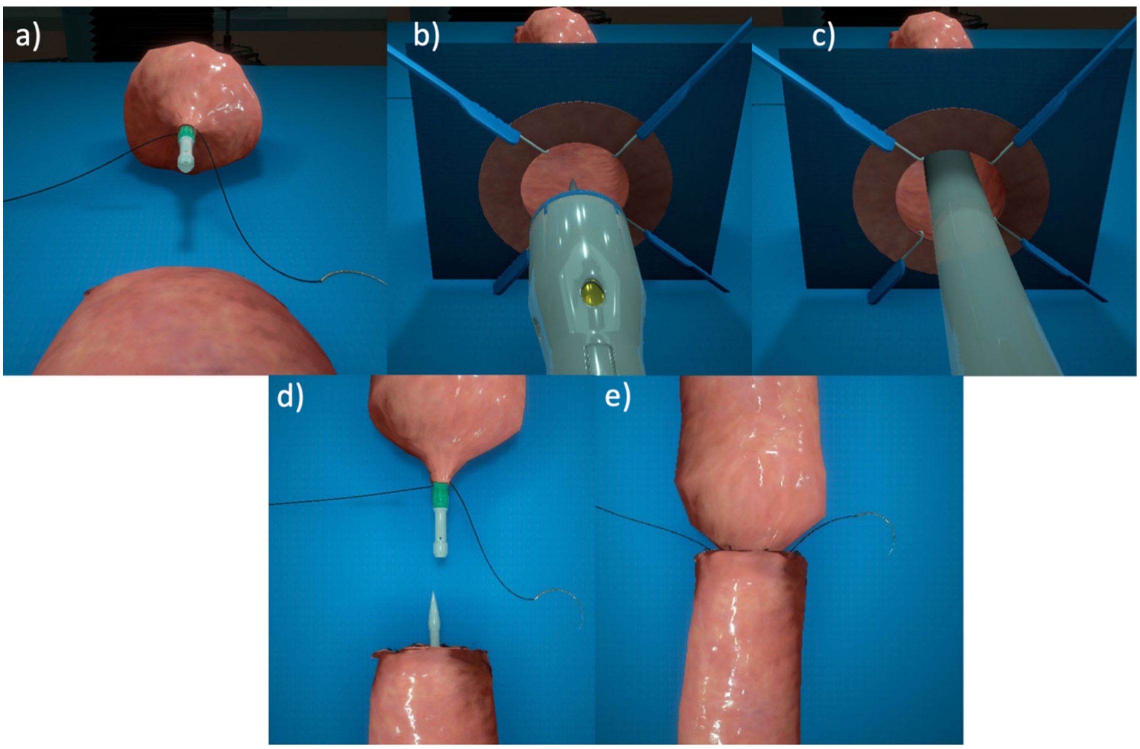

Fig. 5 **a** Adjustment of the anvil, **b** Introduction of circular stapler into the rectal stump, **c** Side view of insertion of the circular stapler, **d** Top view of Insertion of the circular stapler, and **e** Connection of circular stapler to the anvil

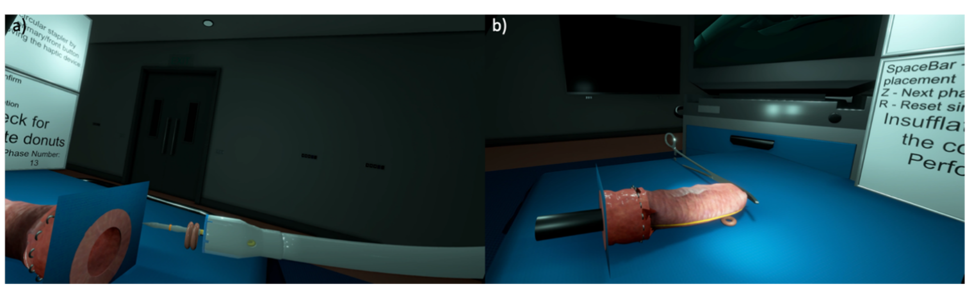

Fig. 6 **a** Inspection of the colonic donuts produced by the firing of the circular stapler and **b** Insufflation of the colon with air to check for leaks

Table 1 Task-specific metrics developed for the assessment of straight coloanal anastomosis

No	Metrics	Score
M1	Transect proximal colon at planned site of anastomosis and hold bowel open	
	Use 2 allis forceps to hold it open	5
	Use 1 allis forceps to hold it open	0
M2	Type of sutures (purse string)	
	Monofilament	5
	Braided Suture	0
M3	Evaluation and inspection of blood flow	
	Ensuring that blood flow to the segment is adequate; pulsatile, bright red confirming the excellent perfusion	5
	Inadequate blood flow; dark blood from the marginal artery	0 (fail)
M4	Technique of placing purse string	
	Equidistance placement of sutures (over and over or Connell): 2 mm apart	5
	Poor placement of sutures/ not full thickness	0
M5	Suture handling for tying purse string	
	Both ends of sutures facing outside	5
	One suture facing outside	0
M6	Insert anvil of stapler into the lumen of the colon	
	Insert Anvil without tearing the purse string	5
	Insert Anvil via allis forceps with tear	0
M7	Secure anvil	
	Solid purse string	5
	Gaps in purse string	0
M8	Inspection of transverse stapler across rectum	
	Fully closed	5
	Inadequate closure	0 (fail)
M9	Introduction of stapler through anal canal	
	Passage up to staple line to flatten top of rectum	5
	Top of rectum is off center or redundant rectum prevents flap surface	0
M10	Stapler motion	
	Gentle insertion until the transvers staple line is flattened	5
	Tears Rectum	0
M11	Stabilizes rectal stapler	
	Maintains stability of the stapler to avoid retraction of post	5
	Failure to advance the stapler and introduce kinks and folds	0
M12	Deploy central trocar/spike	
	At or near the middle of staple line	5
	Too far Anterior or Posterior or Multiple holes	0 (fail)
M13	Reassemble the anvil	
	Audible click is heard when connected to the spike and bowel aligned	5
	Secure connection is failed or twisted bowel	0
M14	Stapler is fully closed	
	Uses gap indicator appropriately to determine sage closure	5
	Fails to use gap indicator	0
M15	Inspection prior to firing	

Table 1 (continued)

No	Metrics	Score
	The anvil and stapler are joined at the staple line	5
	Adjacent tissue is entrapped or mesentery is twisted	0
M16	Safety tab of the stapler	
	Removes safety	5
	Fires through safety	0
M17	Opening stapler after firing	
	Appropriate number of turns for stapler	5
	Inadequate turns	0
M18	Extraction of stapler safely	
	Rotates and removes stapler pushes in to allow anvil to release staple line	5
	Pulls stapler without rotating	0
M19	Assess donuts	
	2 donuts complete	5
	Incomplete donuts	0
M20	Anastomosis check	
	Checking for integrity and leakage (air or betadine)	5
	Not checking	0
M21	Perforation during the procedure	
	Recognizes and Repairs leak	5
	Fails to recognize leak	0 (fail)
M22	Staple line bleeding intervention	
	Complete suture ligation	5
	Incomplete	0 (fail)
M23	Tension of the anastomosis	
	Tension free	5
	Inadequate length and stiff colon	0
M24	Evaluation for bleeding	
	Check bleeding with proctoscope	5
	No proper examination for bleeding	3
General metrics		
M25	Tool handling	
	Smoothness and gentleness in tool handling	5
	Discrete motions in tool handling	3
	Aggressive tool handling	0
M26	Motion	
	Economy of moves	5
	Some unnecessary moves	3
	Unnecessary moves	0
M27	Knowledge of instruments and procedure	
	Knowing all features about instruments and procedure	5
	Having enough knowledge	3
	Deficient knowledge	0
M28	Tasks execution order	
	Completion of tasks executed in order	5
	Completion of tasks executed not in order	0

Table 1 (continued)

No	Metrics	Score
M29	Completion time	
	Under 15 min	5
	15 < 30 min	3
	More than 30 min	0

Table 2 Averages and Welch's t test scores comparing novices and experts based on metric scores

	Purse suture	Trocar	Anvil	Circular stapler	Colonoscope	Total metric score
Average novice score	1.111	4.444	3.889	5.000	3.889	38.333
Average expert score	2.857	5.000	5.000	4.286	5.000	42.143
t value	− 1.5002	− 1.0606	− 1.6036	− 1.0801	− 1.6036	− 3.119
D.O.F	12	8	8	6	8	13
Critical value	1.782	0.16	1.86	2.45	1.86	1.771
p-value	0.0797	0.0741	0.0738	0.3216	0.0738	0.0041

were asked to evaluate the simulator's realism, the usefulness of the force feedback, and to rate the six subscales of NASA-TLX [36] to assess the mental and physical workload experienced during simulator use. To assess the participants' performance, we used two types of metrics: a) performance metrics derived using the HTA and obtained through expert consensus, as described in Sect. "Expert consensus survey" and also listed in Table 1, and b) simulation metrics derived from data collected from the simulator, including device path length, acceleration, velocity, and jerk of the headset and haptic devices. The data collected from the study was analyzed using an unpaired one-tailed Welch's t test to calculate the significance of our results.

The simulation is preceded by a quiz in which participants are required to order the steps of the operation. All operation steps are provided to the participants, who must arrange them correctly to score points. This quiz ensures that participants are familiar with the operation, adding a layer to differentiate between novices and experts. However, the outcome of this quiz did not affect the events in the actual simulation or its metric scores. Once the quiz is complete, the correct steps and their final score are shown to the user, as seen in Fig. 7.

Results

The pre- and post-questionnaire data allowed us to gain anonymous insights into the participants. Performance metrics facilitated an automated and objective assessment. The haptic and VR headset data obtained from the simulator expanded our options for objective evaluation.

Expert consensus survey results

Among the 29 performance metrics evaluated (in Sect. "Expert consensus survey" and Table 1), a mere 1 metric item (3.4%) attained a weighted average between 3 and 4, underscoring the predominantly high importance of the developed metrics. Weighted average scores (Table 3) for the 24 task-specific metrics varied from 3.5 (above neutral) to 5 (of utmost importance), and for the general metrics, the range extended from 3.3 (above neutral) to 4.6 (above important).

The agreed-upon task-specific metrics for assessing straight coloanal anastomosis are categorized into three major procedural groups: 1) bowel transection/creation of purse string around anvil (metrics 1–7), 2) creation of straight coloanal anastomosis (metrics 8–18), and 3) anastomotic assessment (metrics 19–24), with 4) general metrics separately presented as the fourth category.

Within the first set of metrics (bowel transection/creation of purse string around anvil), all seven metrics were rated as very important, with weighted averages ranging from 3.53 (selecting type of sutures for purse strings) to 4.77 (evaluation and inspection of blood flow and securing anvil with purse string). For the creation of straight coloanal anastomosis (metrics 8–18), the weighted average scores ranged from 4.42 (inspection of transverse staple line across rectum) to 4.81 (reassembling the anvil). In the creation of the anastomotic assessment (metrics 19- 24), the weighted average scores ranged from 4.23 (evaluation for bleeding) to 5 (anastomosis check for integrity and leakage). As for the general metrics, the weighted average scores ranged from 1 (type of

Table 3 Weighted average scores for task-specific metrics developed for the assessment of straight coloanal anastomosis

No	Metrics	Agreement (weighted average)
M1	Transect proximal colon at planned site of anastomosis and hold bowel open	4.50
M2	Type of Sutures (Purse String)	3.54
M3	Evaluation and Inspection of blood flow	4.77
M4	Technique of placing purse string	4.31
M5	Suture handling for tying purse string	4.12
M6	Insert anvil of stapler into the lumen of the colon	4.65
M7	Secure anvil	4.77
M8	Inspection of transverse stapler across rectum	4.42
M9	Introduction of stapler through anal canal	4.58
M10	Stapler motion	4.62
M11	Stabilizes rectal stapler	4.47
M12	Deploy central trocar/spike	4.47
M13	Reassemble the anvil	4.81
M14	Stapler is fully closed	4.69
M15	Inspection prior to firing	4.69
M16	Safety tab of the stapler	4.46
M17	Opening stapler after firing	4.50
M18	Extraction of stapler safely	4.50
M19	Assess donuts	4.69
M20	Anastomosis check	5.0
M21	Perforation during the procedure	4.92
M22	Staple line bleeding intervention	4.35
M23	Tension of the anastomosis	4.89
M24	Evaluation for bleeding	4.23
General metrics		
M25	Tool handling	4.35
M26	Motion	4.0
M27	Knowledge of instruments and procedure	4.54
M28	Tasks execution order	4.31
M29	Completion time	3.35

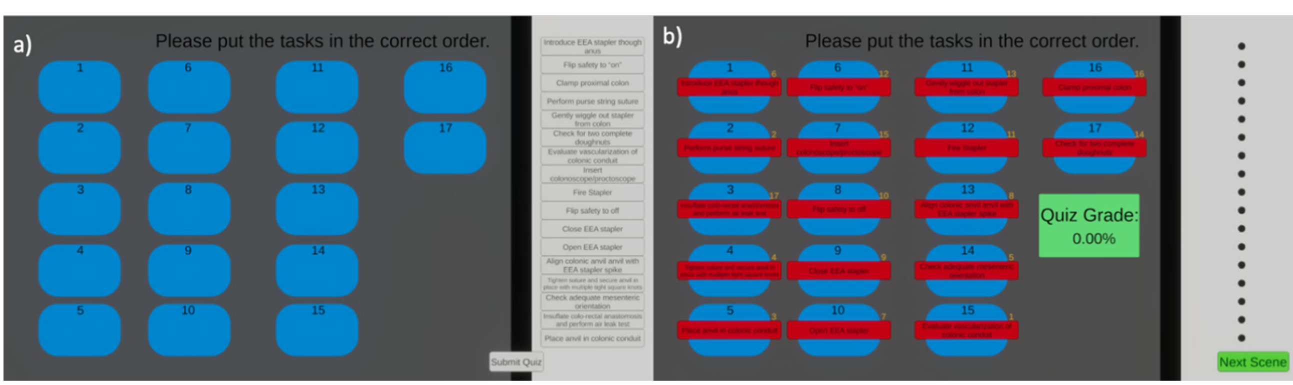

Fig. 7 **a** Quiz module and **b** Completed quiz module. The numbers in orange represent the correct step

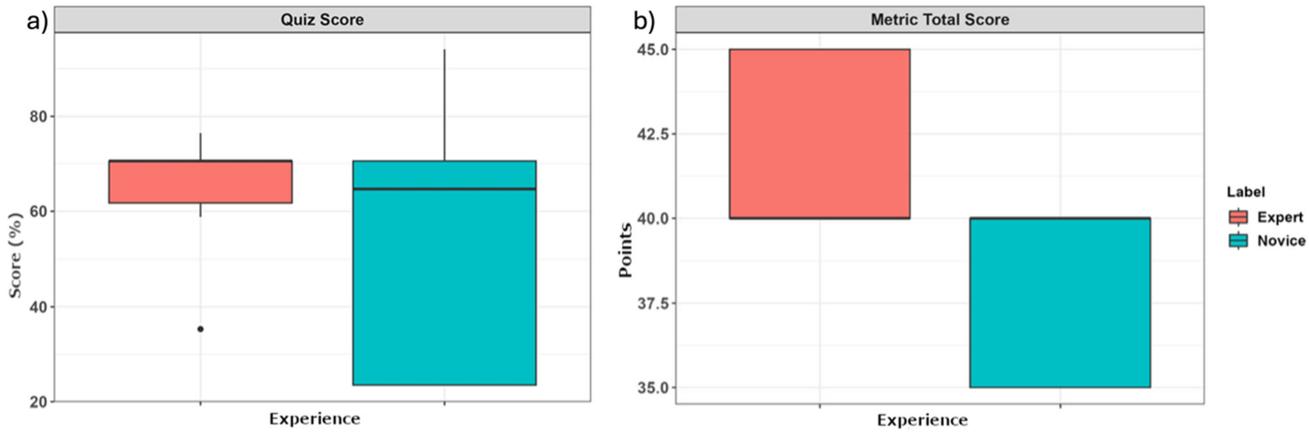

Fig. 8 **a** Box plot of quiz scores comparing novices and experts and **b** Box plot of total metric scores comparing novices and experts

suture, suture handling, and task completion) to 4.53 (knowledge of instruments and procedure).

Performance metric comparison of expert and novice surgeons

We examined the participant's performance on the quiz administered at the start of the simulator. The comparison between experts and novices revealed a disparity in performance, as seen in Fig. 8a. The experts exhibited higher quiz scores on average when compared to the novices. The average score for experts was 63.87%, whereas novices averaged 53.59%.

In our simulation, performance metrics served as the scoring mechanism as seen in Table 1. Scoring was automatically carried out by the simulation according to the metric items. This was done to remove any subjectivity. Participants received 5 points if a task was completed correctly, but they would receive 0 points if completed incorrectly. The simulation involved 21 metric items (M1–M20 and M23), of which nine (M1, M4, M5, M9, M12, M15, M17, M18, M20) were deemed measurable in the simulation by our expert surgeons, amounting to 45 points. The only excluded metrics are M21, M22, M24, and the general metrics. Among these nine metrics, only five returned sets of data that differentiated expert and novice performance:

- (*Purse Suture*) Suture handling for tying Purse Suture (M5).
- (*Trocar*) Deploy central trocar/spike (M12).
- (*Anvil*) Inspection prior to firing (M15).
- (*Circular Stapler*) Extraction of the circular stapler from the anus (M18).
- (*Colonoscope*) Completely inserting the colonoscope into the anus for insufflation to check for integrity and leakage (M20).

Experts outperformed novices by 8.47% for the total metric score, with a novice standard deviation of 2.5 and an expert standard deviation of 2.67. The box plot visualizing the total metric scores for expert and novice groups can be seen in Fig. 8b. Of the five mentioned metrics, novices only performed better in the *Circular Stapler* metric. Experts scored 34.92% better than novices for the *Purse Suture* metric, with a standard deviation of 2.205 for novices and 2.673 for experts. Experts outperformed novices in the *Trocar, Anvil,* and *Colonoscope* tasks by 11.12%, 22.2%, and 22.2%, respectively. The standard deviations for novices were 1.607, 2.205, and 2.205, respectively. The average scores of novice and expert groups for the five mentioned metric items and total metric scores can be seen in Table 2.

Also, Table 2 shows the metrics (purse suture, trocar, anvil, colonoscope, and total metric scores) with differences between expert and novice surgeons using the unpaired one-tailed Welch's t test. Beyond the individual metrics, the total metric score shows a significant difference between experts and novices (alpha = 0.05). The p-value for the t test total metric score comparing experts and novices was 0.0041, showing a significant difference between the groups. The average score for experts was 42.143 out of 45 (93.65%), while novices had an average score of 38.333 out of 45 (85.18%).

Simulation metric comparison of expert and novice surgeons

Along with performance metrics, we used data from the simulator to compare expert and novice groups. In the unity development framework, each unit represents 1 m. However, the scene used for the simulation was scaled by a factor of 60. For clarity, we have used simulated units (su), where 1 simulated unit is equivalent to 1.67 cm. We compared the device path length, acceleration, velocity, and jerk (acceleration per second) of the headset and the haptic devices. The novice

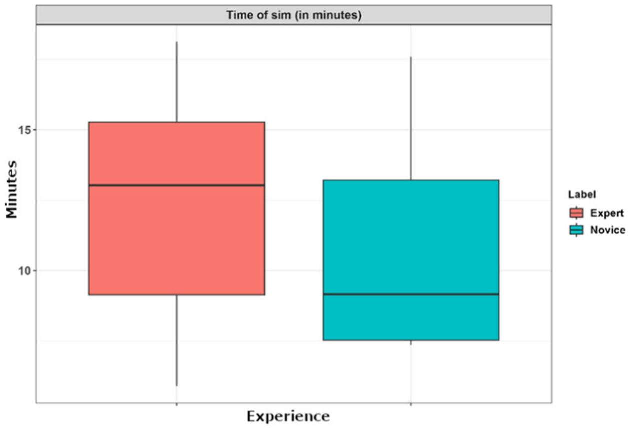

Fig. 9 Box plots of time taken to complete the simulation comparing novice and expert groups

group (11.046 min) completed the simulation 10.49% faster than the expert group (12.27 min). The box plot in Fig. 9 depicts the average time to complete the simulation for the novice and expert groups.

We have aggregated the left and right haptic device data for reporting, as surgeons must use both hands equally in colorectal surgery. For novice and expert groups, the performance metrics related to haptic device usage specifically focus on aggregate jerk, aggregate acceleration, aggregate device path length, and aggregate velocity. Regarding aggregate jerk, experts demonstrated a slightly lower jerk (2.321 su/s^3) than novices (2.406 su/s^3), indicating smoother and more controlled movements. A similar trend is observed in aggregate acceleration, where experts exhibited a lower acceleration (1.332 su/s^2) in contrast to novices (1.404 su/s^2), implying a more refined handling of the device's speed changes. However, aggregate device path length showed that novices (1497.072 su) moved their haptic devices less compared to experts (2555.22 su). Additionally, experts achieved a lower score (0.901 su/s) in aggregate velocity than novices (0.985 su/s), suggesting that experts maintained a steadier velocity while interacting with the haptic device. The box plots for the simulation performance metrics, aggregate jerk, aggregate acceleration, aggregate device path length, and aggregate velocity can be seen in Fig. 10.

Headset data showed opposite results compared to the haptic device data. Regarding headset average velocity per second, experts achieved a significantly higher value of 932.309 su/s compared to novices' 645.595 su/s, indicating faster movements. This increased velocity is complemented by a higher average acceleration among experts (1.032 su/s^2) compared to novices (0.828 su/s^2). Moreover, for the jerk per second, experts demonstrated a slightly higher jerk (1.743 su/s^3) than novices (1.405 su/s^3). Figure 11 shows the box plots for the simulation performance metrics: headset jerk, acceleration, velocity, and path length. Also, the NASA-TLX scores indicating mental and physical workload during the simulation were 31.14 for the experts and 31.22 for the novices.

Discussion

The results of this study show a significant difference between novices and experts in completing the SCA simulation. This can be seen through t testing metric scores and analyzing average positional data of all devices used during the simulation. To test not just technical skills, the simulation also tasked participants with a quiz about the operation. From the data presented in Sect. "Performance metric comparison of expert and novice surgeons", experts scored 10.28% more on the quiz than novices on average. Experts were consistent in their scores, while novices varied greatly in the range of their scores. The only differences in the scores for the expert group came from the incorrect order of direct sequential steps. These findings can imply that experts are familiar with the steps but do not remember minor differences between them, giving them a consistent idea of the order in which they would perform the operation. In terms of the NASA-TLX scores, the small difference of 0.08 between the two suggests that in this specific procedure, the workload perceived by the expert and novice participants is almost identical. This means that the task is equally challenging for both experts and novices, and that the task design does not significantly advantage or disadvantage either group in terms of perceived workload.

As shown in Table 1, we initially considered 24 specific and five general metrics to assess the SCA procedure. The general metrics were evaluated through simulation metrics. However, because blood flow was not simulated, metrics M21, M22, and M24 were excluded, reducing the total number of metrics used in the simulator to 21 (M1–M20 and M23). Of these 21 metrics, only nine (M1, M4, M5, M9, M12, M15, M17, M18, and M20) were ultimately used for assessment, as simulating the remaining metrics were either too costly (e.g., rectal tears, twisted bowel, etc.) or would have resulted in automatic full scores due to the simulator's progression. From the averages of the individual metric scores, it can be surmised that experts, on average, scored better on metrics than novices. This can be expounded upon by examining the t tests showing a significant difference in select metrics. From the individual tracked metrics, three showed differences between novices and experts. They were the *Purse Suture* with a p-value of 0.0797, *Anvil* with a p-value of 0.0738, and *Colonoscope* with a p-value of 0.0738. In terms of averages, experts scored better than the novices in these metrics by 1.746 points for purse suturing, 1.111 points for anvil retraction, and 1.111 for insertion of colonoscope.

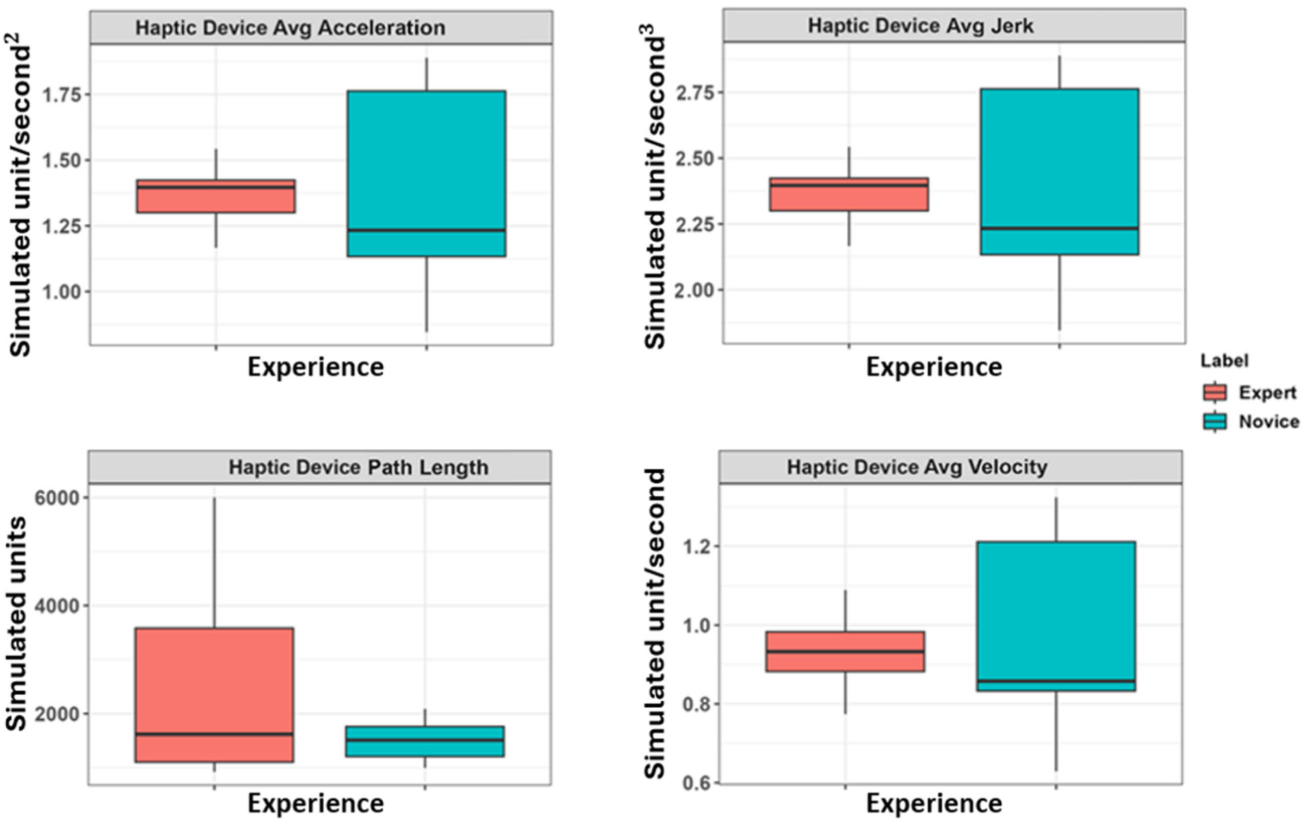

Fig. 10 Box plots of haptic device simulation performance metrics for novice and expert groups

According to the haptic device positional data, experts are more likely to spend time performing tasks with a more extensive range of motion but smoother movements when compared to novices. Expert surgeons have developed muscle memory through repeated practice of the procedure [37]. To validate this claim, we averaged the experts' and novices' velocity, acceleration, jerk, and path length. The results show that experts better control their hand movements and limit jerky movements that are detrimental in highly skilled surgical environments. Also, their movements are more consistent with other experts, while novices have a larger variance across their movements. This consistency and smoothness show a level of skill not found with novices. However, on the other hand, novices seem to understand and function smoother than experts according to the VR headset's motion. We believe that this difference between novices and experts comes from experience with both simulators and procedures. Experts are much more experienced when it comes to testing simulators and performing the simulated surgeries, leading to them having preconceived notions and muscle memory differing from the new simulators that involve VR. On the other hand, novices are much less ingrained with experience and are more likely to adapt to a new method of simulation without having conflicting conceptions or techniques. Overall, the performance metrics— *Purse Suture, Trocar, Anvil, Circular Stapler, Colonoscope, and Total Metric Score*—and the simulation metrics—VR headset motion and haptic device data—were able to show differences between the expert and novice groups.

In the post-questionnaire, the participants scored the realism of the anatomy in the simulator as a three out of five. Only one participant scored the realism a 1 (low realism), four participants scored the realism a 4, while the rest were clustered around 3. The experts scored the realism higher (3.3) compared to novices (2.7). These results were expected since experts have been exposed to a higher number of surgical simulators, while the novices lack experience or reference points like the experts. With more simulators to compare to, experts are more likely to score with other simulations in mind.

Conclusion

In conclusion, we designed and developed a VR-based SCA simulator with a performance metric that can differentiate novice and expert performance. As part of the study, we conducted a user study with experts ($n = 7$) and novices ($n = 9$). In our study, experts performed better than novices when

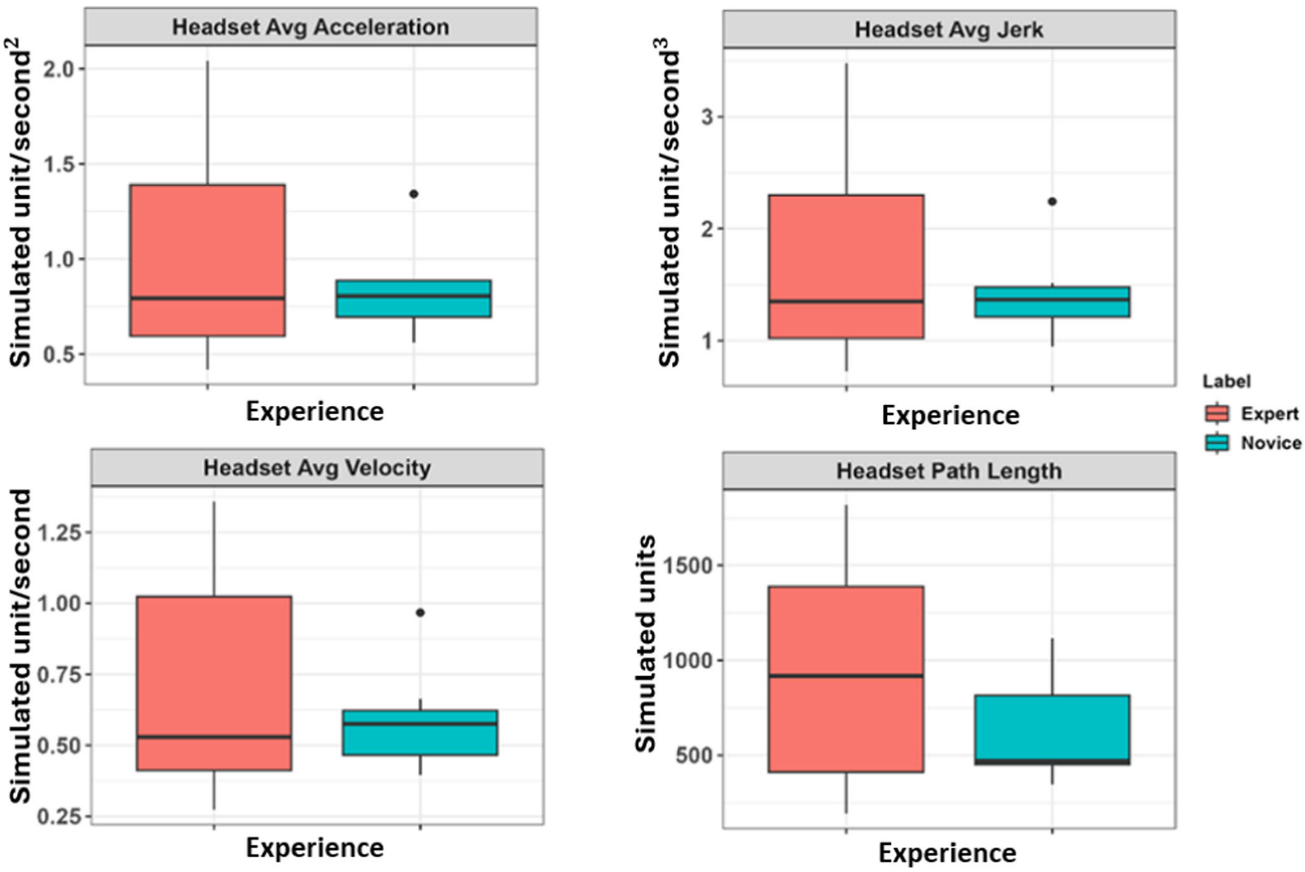

Fig. 11 Box plots of headset simulation performance metrics box plots for novice and expert groups

completing tasks for the surgery. Experts scored, on average, 3.80 points (8.5%) better than novices in our performance metrics.

Additionally, experts demonstrated superior knowledge through a 10.28% higher quiz score, indicating their deep understanding of the surgical procedure. Notably, experts exhibited smoother hand movements, evidenced by a 3.6% less jerk, a 5.2% less acceleration, and an 8.6% less velocity compared to novices. These findings underscore the enhanced precision and skill exhibited by experts, highlighting the potential of our simulator to bridge the proficiency gap between novices and experts.

As part of our study, we conducted an Expert Consensus Survey ($n = 27$). Using a 5-Point Likert Scale, participants rated the significance of 29 performance metrics. Only 1 metric (3.4%) had a weighted average between 3 and 4, highlighting the overall high importance of the developed metrics. Task-specific metrics ranged from 3.5 to 5, emphasizing their significance, while general metrics had a range of 3.3 to 4.6. Notably, within specific metric sets, such as bowel transection and the creation of purse string, all seven metrics were rated as very important. The study provides a comprehensive understanding of the perceived importance of various surgical metrics among experienced colorectal surgeons. The alignment between the expert opinions and the performance metrics used in the simulator enhances our work's relevance.

Our research establishes the foundation for an SCA training simulator with the potential for users to analyze their skills with deliberate metric scores and quiz grading. To the best of our knowledge, this is the first SCA simulator that uses VR for immersion, haptic feedback to feel physical sensations, and utilizes objective performance metrics to differentiate between expert and novice performances. Knowing where you are inadequate allows for a more specific approach and training routine to speed up the learning process and expedite technical skill enhancement. Furthermore, it also shows the user's their smoothness in hand movements, allowing for objective scrutiny of their motions. The objective metrics can improve the smoothness of hand movements and deliberate feedback could speed learning. Any user can obtain objective scores and focus on improving their technique, thereby paving the way for enhanced patient outcomes and safety standards.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11548-024-03291-z.

Acknowledgements This project was supported by grants from the National Institutes of Health (NIH)/NIBIB R01EB025241, R01EB033674, R01EB032820, and R01EB005807.

Declarations

Conflict of interest George Westergaard, Jacob Barker, Drs. Alexis Desir, Ganesh Sankaranarayanan, Tansel Halic, Shruti Hegde, Amr Al Abbas, Javier Salgado Pogacnik, James W. Fleshman, Doga Demirel, and Suvranu De have no conflicts of interest or financial ties to disclose.

Ethical approval This study was approved by the University of Texas Southwestern Institutional Review Board (IRB # STU-2021–0202).

Informed consent Informed consent was obtained from all participants included in the study.

References

1. de Montbrun S, Roberts PL, Satterthwaite L, MacRae H (2016) Implementing and evaluating a national certification technical skills examination. Ann Surg 264:1–6
2. De Montbrun SL, Roberts PL, Lowry AC, Ault GT, Burnstein MJ, Cataldo PA, Dozois EJ, Dunn GD, Fleshman J, Isenberg GA (2013) A novel approach to assessing technical competence of colorectal surgery residents: the development and evaluation of the colorectal objective structured assessment of technical skill (COSATS). Ann Surg 258:1001–1006
3. Caulfield H, Hyman NH (2013) Anastomotic leak after low anterior resection: a spectrum of clinical entities. JAMA Surg 148:177–182
4. Detry RJ, Kartheuser A, Delriviere L, Saba J, Kestens PJ (1995) Use of the circular stapler in 1000 consecutive colorectal anastomoses: experience of one surgical team. Surgery 117:140–145
5. Heald RJ, Leicester RJ (1981) The low stapled anastomosis. Br J Surg 68:333–337
6. Kotby MN, Wahba HA, Kamal E, El-Makhzangy AMN, Bahaa N (2012) Animal model for training and improvement of the surgical skills in endolaryngeal microsurgery. J Voice 26:351–357
7. Denadai R, Saad-Hossne R, Todelo AP, Kirylko L, Souto LRM (2014) Low-fidelity bench models for basic surgical skills training during undergraduate medical education. Rev Col Bras Cir 41:137–145
8. Sararit N, Haddawy P, Suebnukarn S (2018) Effectiveness of a low-cost VR simulator for emergency management training in dental surgery. In: 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, pp 1–6
9. Van Nguyen D, Lakhal SB, Chellali A (2015) Preliminary evaluation of a virtual needle insertion training system. In: 2015 IEEE Virtual Reality (VR). IEEE, pp 247–248
10. Wu J-X, Chen G-C, Chang C-W, Lin C-H (2016) Development of virtual-reality simulator system for minimally invasive surgery (MIS) using fractional-order vascular access. In: 2016 SAI Computing Conference (SAI). IEEE, pp 1257–1259
11. Vaughan N, Dubey VN, Wainwright TW, Middleton RG (2015) Does virtual-reality training on orthopaedic simulators improve performance in the operating room? In: 2015 Science and Information Conference (SAI). IEEE, pp 51–54
12. Salleh R, Loureiro R, Caldwell DG (2004) Haptic feedback for VR-based minimally invasive surgical (MIS) training. Meas Control 37:174–182
13. Chemlal S, Rechowicz KJ, Obeid MF, McKenzie FD (2014) Poster: Modeling insertion point for general purpose haptic device simulations for minimally invasive surgeries. In: 2014 IEEE Symposium on 3D User Interfaces (3DUI). IEEE, pp 129–130
14. Kannangara SM, Fernando E, Nanayakkara ND, Kumarage SK (2016) Evaluating effects of haptic feedback in Virtual Reality simulators for laparoscopic skill development. In: 2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob). IEEE, pp 396–400
15. Alvarez-Lopez F, Maina MF, Saigí-Rubió F (2020) Use of a low-cost portable 3d virtual reality gesture-mediated simulator for training and learning basic psychomotor skills in minimally invasive surgery: development and content validity study. J Med Internet Res 22:e17491
16. Demirel D, Yu A, Halic T, Sankaranarayanan G, Ryason A, Spindler D, Butler KL, Cao C, Petrusa E, Molina M, Jones D, De S, Demoya M, Jones S (2016) Virtual airway skills trainer (VAST) simulator. Stud Health Technol Inform 220:91–97
17. Demirel D, Butler KL, Halic T, Sankaranarayanan G, Spindler D, Cao C, Petrusa E, Molina M, Jones DB, De S (2016) A hierarchical task analysis of cricothyroidotomy procedure for a virtual airway skills trainer simulator. Am J Surg 212:475–484
18. Qi D, Petrusa E, Kruger U, Milef N, Abu-Nuwar MR, Haque M, Lim R, Jones DB, Turkseven M, Demirel D, Halic T, De S, Saillant N (2020) Surgeons with five or more actual cricothyrotomies perform significantly better on a VR simulator. J Surg Res 252:247–254. https://doi.org/10.1016/j.jss.2020.03.021
19. Qian K, Jiang T, Wang M, Yang X, Zhang J (2016) Energized soft tissue dissection in surgery simulation. Comput Anim Virt Worlds 27:280–289
20. Shi Y, Liu M, Xiong Y, Cai C, Tan K, Pan X (2015) The simulation of delineation and splitting in virtual liver surgery. In: 2015 International Conference on Virtual Reality and Visualization (ICVRV). IEEE, pp 264–268
21. Li R, Lou Y (2019) Research and development of the regulation training system of minimally invasive knee surgery based on VR technology. In: 2019 14th International Conference on Computer Science & Education (ICCSE). IEEE, pp 161–164
22. Gao B, Shang L, Cai X, Jiang Y, Yang S (2019) Blood Flow Simulation of Virtual Simulation System for Vascular Interventional Surgery. In: 2019 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, pp 2245–2250
23. Wijewickrema S, Copson B, Zhou Y, Ma X, Briggs R, Bailey J, Kennedy G, O'Leary S (2017) Design and evaluation of a virtual reality simulation module for training advanced temporal bone surgery. In: 2017 IEEE 30th international symposium on computer-based medical systems (CBMS). IEEE, pp 7–12
24. Sommer GM, Broschewitz J, Huppert S, Sommer CG, Jahn N, Jansen-Winkeln B, Gockel I, Hau H-M (2021) The role of virtual reality simulation in surgical training in the light of COVID-19 pandemic: visual spatial ability as a predictor for improved surgical performance: a randomized trial. Medicine 100:
25. Hirano M, Sakurada M, Furuya S (2020) Overcoming the ceiling effects of experts' motor expertise through active haptic training. Science advances 6:eabd2558
26. Macklin M, Müller M, Chentanez N (2016) XPBD: position-based simulation of compliant constrained dynamics. In: Proceedings of the 9th International Conference on Motion in Games. pp 49–54
27. Sankaranarayanan G, Parker LM, Khan A, Dials J, Demirel D, Halic T, Crawford A, Kruger U, De S, Fleshman JW (2023) Objective metrics for hand-sewn bowel anastomoses can differentiate novice from expert surgeons. Surg Endosc 37:1282–1292. https://doi.org/10.1007/s00464-022-09584-1

28. Dials J, Demirel D, Sanchez-Arias R, Halic T, Kruger U, De S, Gromski MA (2023) Skill-level classification and performance evaluation for endoscopic sleeve gastroplasty. Surg Endosc 37:4754–4765. https://doi.org/10.1007/s00464-023-09955-2
29. Sankaranarayanan G, Parker LM, Jacinto K, Demirel D, Halic T, De S, Fleshman JW (2022) Development and validation of task-specific metrics for the assessment of linear stapler-based small bowel anastomosis. J Am Coll Surg 235:881. https://doi.org/10.1097/XCS.0000000000000389
30. Demirel D, Palmer B, Sundberg G, Karaman B, Halic T, Kockara S, Kockara N, Rogers ME, Ahmadi S (2022) Scoring metrics for assessing skills in arthroscopic rotator cuff repair: performance comparison study of novice and expert surgeons. Int J CARS 17:1823–1835. https://doi.org/10.1007/s11548-022-02683-3
31. Hegde S, Al Abbas A, Sankaranarayanan G, Dimirel D, Halic T, Pogacnik JS, De S, Fleshman JW (2022) development and expert consensus on performance metrics for automated assessment of stapled straight coloanal anastomosis. in: diseases of the colon & rectum. Lippincott Williams & Wilkins Two Commerce Sq, 2001 Market st, Philadelphia …, pp 129–129
32. Demirel D, Yu A, Cooper-Baer S, Dendukuri A, Halic T, Kockara S, Kockara N, Ahmadi S (2017) A hierarchical task analysis of shoulder arthroscopy for a virtual arthroscopic tear diagnosis and evaluation platform (VATDEP). Int J Med Robot Comput Assist Surg 13:e1799. https://doi.org/10.1002/rcs.1799
33. Dials J, Demirel D, Halic T et al (2022) Hierarchical task analysis of endoscopic sleeve gastroplasty. Surg Endosc 36:5167–5182
34. Cetinsaya B, Gromski MA, Lee S, Xia Z, Demirel D, Halic T, Bayrak C, Jackson C, De S, Hegde S, Cohen J, Sawhney M, Stavropoulos SN, Jones DB (2019) A task and performance analysis of endoscopic submucosal dissection (ESD) surgery. Surg Endosc 33:592–606. https://doi.org/10.1007/s00464-018-6379-6
35. Cohen PR, Martinelli PT, Schulze KE, Nelson BR (2007) The purse-string suture revisited: a useful technique for the closure of cutaneous surgical wounds. Int J Dermatol 46:341–347. https://doi.org/10.1111/j.1365-4632.2007.03204.x
36. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Human mental workload. Elsevier, pp 139–183
37. Goff BA (2008) Changing the paradigm in surgical education. Obstet Gynecol 112:328–332

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.