

Original Article

Open Access



Competency-Based Assessment of Robotic Surgery Skills - Phase 2 (CARS 2.0): A Global Survey Study with Blinded Video Review of Surgical Proficiency On Behalf of TROGSS - The Robotic Global Surgical Society

Aman Goyal^{1,2} , Abhirami Babu³, Susana Acosta⁴, Yeisson Rivero-Moreno^{5,6} , Christian A. Macias^{7,8,9} , Vanessa Pamela Salolin Vargas¹⁰ , Jason Park¹¹ , Elisa Rodriguez¹¹, Kimberly Ibarra¹¹, Michail Koutentakis^{12,13}, Clotilde Fuentes-Orozco¹⁴ , Andrea Garcia¹⁴ , Kathia Dayana Morfin Meza¹⁴ , Alejandro González-Ojeda¹⁵ , Luis Osvaldo Suárez-Carreón^{16,17} , Adel Abou-Mrad¹⁸ , Luigi Marano^{19,20,21,22} , Rodolfo J. Oviedo^{11,23,24} ; On Behalf of TROGSS, Collaborative Research Consortium

¹Department of General Surgery, Mahatma Gandhi Medical College and Research Institute, Puducherry 607402, India.

²Department of Surgery, Adesh Institute of Medical Sciences and Research, Bathinda 151109, India.

³College of Medicine, Our Lady of Fatima University, Valenzuela City 1440, Philippines.

⁴Department of Surgery, Universidad Central de Venezuela, Caracas 1020, Venezuela.

⁵Department of Surgery, Montefiore Medical Center, New York, NY 10467, USA.

⁶Department of Surgery, Universidad de Oriente, Núcleo Anzoátegui, Barcelona 6001, Venezuela.

⁷Health Science Department, Hillsborough College, Tampa, FL 33614, USA.

⁸School of Medicine, Universidad Catolica de Santiago de Guayaquil, Guayaquil 090101, Ecuador.

⁹Surgical Robotic Research Division, Center for Space Systems (C-SET), Lima 15001, Peru.

¹⁰Facultad de Medicina, Universidad Westhill, Ciudad de México 05610, México.

¹¹College of Osteopathic Medicine, Sam Houston State University, Conroe, TX 77304, USA.

¹²Doctoral School, Medical University of Warsaw, Warsaw 02-091, Poland.

¹³Department of Experimental and Clinical Pharmacology, Medical University of Warsaw, Center for Preclinical Research and Technology (CEPT), Warsaw 02-091, Poland.

¹⁴Unidad de Investigación Biomédica 02, Hospital de Especialidades, Centro Médico Nacional de Occidente, Guadalajara 44100, México.

¹⁵Facultad de Medicina, Universidad de Colima, Colima 28040, México.

¹⁶Department of Bariatric Surgery, UMAE Hospital de Especialidades del Centro Medico Nacional de Occidente, Guadalajara 44349, Mexico.

¹⁷Department of Surgery, Universidad de Guadalajara, Guadalajara 44340, Mexico.

¹⁸Department of Surgery, Centre Hospitalier Universitaire d'Orléans, Orléans 45000, France.

¹⁹Department of Medicine, Academy of Applied Medical and Social Sciences-AMiSNS (Akademia Medycznych I Społecznych Nauk Stosowanych), Elbląg 52-300, Poland.

²⁰Department of General Surgery and Surgical Oncology, "Saint Wojciech" Hospital, "Nicolaus Copernicus" Health Center, Gdańsk 80-000, Poland.

²¹Department of Surgery, Dnipro State Medical University, Dnipro 49044, Ukraine.

²²Department of Medicine, Surgery, and Neurosciences, University of Siena, Siena 53100, Italy.

²³Department of Surgery, Nacogdoches Medical Center, Nacogdoches, TX 75965, USA.

²⁴Department of Surgery, Tilman J. Fertitta Family College of Medicine, University of Houston, Houston, TX 77001, USA.

Correspondence to: Dr. Rodolfo J. Oviedo, Department of Surgery, Tilman J. Fertitta Family College of Medicine, University of



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



Houston, Houston, TX 77001, USA. E-mail: roviado3@central.uh.edu

How to cite this article: Goyal A, Babu A, Acosta S, Rivero-Moreno Y, Macias CA, Vargas VPS, Park J, Rodriguez E, Ibarra K, Koutentakis M, Fuentes-Orozco C, Garcia A, Meza KDM, González-Ojeda A, Suárez-Carreón LO, Abou-Mrad A, Marano L, Oviedo RJ; On Behalf Of TROGSS, Collaborative Research Consortium. Competency-Based Assessment of Robotic Surgery Skills - Phase 2 (CARS 2.0): A Global Survey Study with Blinded Video Review of Surgical Proficiency On Behalf of TROGSS - The Robotic Global Surgical Society. *Mini-invasive Surg.* 2025;9:34. <https://dx.doi.org/10.20517/2574-1225.2025.66>

Received: 11 May 2025 **First Decision:** 14 Aug 2025 **Revised:** 31 Aug 2025 **Accepted:** 19 Sep 2025 **Published:** 28 Oct 2025

Academic Editor: Fernando A. M. Herbella **Copy Editor:** Xing-Yue Zhang **Production Editor:** Xing-Yue Zhang

Abstract

Aim: The Competency-Based Assessment of Robotic Surgery Skills (CARS) scale was developed as a novel approach to assess robotic surgery (RS) skills through 10 relevant RS competencies. CARS 2.0 aimed to expand on the findings of CARS by conducting a global survey study in which participants graded a blinded, edited surgical video with the CARS scale for 7 competencies measurable via video.

Methods: CARS 2.0 is being conducted globally, including participants across medical specialties and training stages, from medical students to attending/consultant surgeons. Participants evaluated a blinded, edited surgical video using the CARS scale via an anonymous Google form, focusing on 7 video-measurable competencies.

Results: A total of 320 responses were collected over 3 months, including 125 (39.06%) attending/consultant surgeons, 98 (30.6%) surgical specialty postgraduate trainees, 96 (30%) medical students, and 1 pre-medical student. ANOVA (analysis of variance) analysis showed that the operator scores increased with the evaluators' level of experience, reaching statistical significance across all 7 competency categories. Spearman's correlation indicated moderate associations between participants' surgical experience and proficiency ($\rho = 0.314$, $P < 0.001$), as well as between their comfort with the CARS scale and proficiency ($\rho = 0.337$, $P < 0.001$). Regression analysis demonstrated that robotic stapler use and camera handling were predictors of higher CARS scores based on participants' experience.

Conclusion: CARS represents a first step toward establishing competency-based assessment of RS performance independent of specific surgical procedures. Its integration into surgical training programs can facilitate trainees' attainment of RS competency. Longitudinal studies could further validate its effectiveness at improving surgical training with its implementation into training curricula.

Keywords: Robotic surgery, competency-based assessment, surgical education, CARS scale, global survey study

INTRODUCTION

The evolution of surgical robotic technology represents a remarkable journey marked by innovation and collaboration between engineering and medical fields, driven by the goal of enhancing surgical precision and patient outcomes. The origins of surgical robotics date back to the early 1980s. The Arthropod, developed by Davies and Simmons in 1983, was one of the earliest innovative robotic systems designed for orthopedic surgery and marked a significant milestone in medical robotics. The Arthropod demonstrated the potential of robotic assistance in surgical procedures, paving the way for future advancements in the field^[1]. This was followed by the Programmable Universal Machine for Assembly (PUMA) 560 robot, an industrial robot developed by Unimation (Westinghouse Electric, Pittsburgh, PA) in collaboration with the National Aeronautics and Space Administration (NASA), which was employed in 1985 to control surgical tools for stereotactic brain biopsies and the resection of thalamic astrocytomas^[2-4].

In the late 1990s, the da Vinci Surgical System, developed by Intuitive Surgical (CA, USA), marked a pivotal milestone in minimally invasive surgery. Introduced as a console for surgeons to remotely control robotic arms connected to the patient, it provided a high-quality, magnified, three-dimensional view for navigating anatomy^[5]. This system revolutionized procedures, particularly in gastrointestinal surgery, by enabling precise, minimally invasive approaches that improved surgical outcomes and accelerated patient recovery^[6]. The widespread adoption of platforms, such as the da Vinci system, has led to exponential growth in robotic procedures, particularly in general surgery, urology, gynecology, and oncology^[7,8]. With this technological advancement, however, comes an equally urgent need for standardized, competency-based methods to assess surgical proficiency across diverse clinical and educational environments.

Although several assessment tools exist, including the Global Evaluative Assessment of Robotic Skills (GEARS), Robotic Objective Structured Assessment of Technical Skills (R-OSATS), and Robotic Ottawa Surgical Competency Operating Room Evaluation (RO-SCORE), these scales have limitations in their construct validity, procedural specificity, or reliance on simulation-based evaluation^[9-11]. Among these, GEARS remains one of the most widely used and validated tools, showing strong internal consistency and inter-rater reliability across specialties^[12]. ARCS, developed to evaluate console-specific skills, also demonstrated strong construct validity but was primarily tested in simulated environments, limiting its application to real operative settings^[13]. Similarly, while R-OSATS has demonstrated high inter- and intra-rater reliability in simulation and dry-lab settings, it lacks validation in real-time or video-based general surgery evaluations^[14].

A recent review by Kutana *et al.*^[15] emphasized the growing need for objective, scalable, and globally applicable surgical assessment tools, particularly as remote education and automated systems become more integrated into surgical training.

In response to this gap, the Competency-Based Assessment of Robotic Surgery Skills (CARS) scale was developed to offer a structured, multi-domain rubric tailored to general surgery and grounded in real-case surgical observation. The CARS scale was introduced by Elzein *et al.*^[16], and its initial validation demonstrated a promising ability to differentiate surgical skill levels among general surgery residents during in-person evaluation. However, further research is needed to assess the CARS scale's acceptability, usability, and applicability in remote or video-based assessments over a broad range of general surgery procedures and evaluator backgrounds.

This study aims to evaluate global perceptions and the feasibility of using the CARS scale for video-based assessments of robotic surgical skills. By engaging a broad cohort of evaluators across varying levels of surgical training and specialties, this research explores the comfort, perceived reliability, and effectiveness of the CARS scale in modern competency-based surgical education.

METHODS

Study design

This study adopted a descriptive, cross-sectional survey design, conducted globally via an online platform using Google Forms. It was anchored in the previously validated CARS scale developed by Elzein *et al.*^[16]. The primary objective was to assess the feasibility, user comfort, and perceived effectiveness of the CARS scale as a tool for evaluating robotic surgical competence in a standardized, competency-based framework.

Participants and recruitment

Participants were recruited using a multimodal strategy, including direct email invitations and outreach via professional surgical and medical community platforms on social media. The survey was designed to capture responses from a broad spectrum of participants, including attending surgeons, surgical trainees, medical students, and pre-medical students, to ensure a comprehensive evaluation of the scale's applicability across varying levels of training and experience. All participants were required to provide electronic informed consent before participating in the study. Institutional Review Board (IRB) approval was not required, as the study involved the evaluation of a de-identified, edited surgical video. Informed consent was implied by voluntary participation, with submission of de-identified responses constituting agreement to participate.

Video intervention and evaluation

Prior to completing the survey, participants were asked to view a standardized, blinded, and randomized video lasting 4 min and 46 s, depicting robotic conversion of vertical banded gastroplasty to Roux-en-Y gastric bypass (RYGB), with partial gastrectomy, performed by both attending surgeons/consultants and postgraduate residents. The video could be viewed multiple times to ensure a thorough assessment. The surgical procedure demonstrated in the video included extensive adhesiolysis, identification of vertical banded gastroplasty anatomy, internal liver retraction using an absorbable barbed suspension stitch, mobilization of the greater and lesser curvatures, entry into the lesser sac, creation of a gastric pouch over a bougie, and an omega loop technique utilizing a 50 cm antecolic jejunal loop. This was followed by linear-stapled gastrojejunostomy, division of the omega loop to form the Roux and biliopancreatic limbs, creation of a linear-stapled jejunojejunostomy, mesenteric defect closure, and an endoscopic anastomotic leak test. The video was presented in high definition to ensure optimal visualization of procedural steps and allow for accurate assessment by respondents [[Supplementary Video 1](#)].

Assessment tool (CARS scale)

The survey incorporated the CARS scale to assess various domains of robotic surgical performance. The tool prompts raters to evaluate specific surgical competencies on a five-point Likert scale (1 = poor, 5 = excellent), with an additional "cannot be assessed" option. The domains included blunt and sharp tissue dissection, tissue handling and retraction, robotic stapler use, arm exchange and instrument swapping, endoscope utilization, intracorporeal suturing and knot tying, and wrist articulation. Each domain was assessed with distinct evaluative criteria. For example, tissue dissection was evaluated based on finesse, anatomical respect, consistency with natural or re-operative planes, and the use of both gross and fine techniques. Tissue handling and retraction were assessed for dynamic manipulation and entry into the correct anatomical planes. Robotic stapler use was evaluated for angulation, injury prevention, and SmartClamp feedback utilization, while arm exchange focused on exposure management and coordinated use of robotic arms. Endoscope utilization was evaluated for camera angles, depth perception, and application of near-infrared fluorescence imaging (Firefly). Suturing and knot tying were assessed for skill efficiency, suture material management, tissue respect, and needle safety. Additional aspects such as port placement, docking, and ergonomics were included in the scale but rated as "cannot be assessed" due to video limitations. This tool enabled participants to simulate the role of an evaluator based on the video demonstration.

Survey variables

Participants were asked to indicate their previous experience with surgical robotic platforms and their current level of surgical training or practice, categorized as consultant/attending surgeon (> 10 years, 5-10 years, < 5 years), surgical trainee, medical student, or pre-medical student. Consultants and trainees were also asked to specify their primary specialty and any relevant subspecialties. The main section of the survey

aimed to assess participants' comfort in using the CARS scale. Specifically, they were asked whether they would feel comfortable using the scale to evaluate a trainee based on video observation, whether they would be comfortable being assessed using the scale if feedback were provided, and whether they could confidently apply the scale to the specific video provided in the survey.

Statistical analysis

Data were analyzed using descriptive and inferential statistical methods with IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY, USA). Means and standard deviations were calculated for each CARS scale domain to summarize participant evaluations. Comparative analyses using *t*-tests or one-way analysis of variance (ANOVA) were employed to explore differences in mean CARS scores across groups stratified by surgical experience level and self-reported proficiency. Additionally, Spearman's rank correlation tests were conducted to examine associations between surgical experience, robotic platform usage, and self-reported proficiency.

To identify predictors of higher CARS scores, multiple regression analysis was performed with independent variables including surgical experience, prior use of a robotic platform, and self-reported proficiency. Factor analysis was conducted to determine latent constructs underlying the CARS scale domains, which may offer insights into broader categories of robotic surgical competence. Furthermore, chi-square tests were utilized to analyze the association between categorical variables, such as participant comfort with the CARS scale and reported proficiency levels.

Output and interpretation

Statistical results were presented using summary tables displaying mean scores and standard deviations for each CARS domain, along with comparative tables illustrating differences across subgroups. Graphical outputs included bar charts representing distribution patterns and scatter plots to visualize correlation trends. Interpretive analysis focused on identifying patterns in participant responses, highlighting statistically significant relationships, and discussing the implications of findings in the context of competency-based evaluation in robotic surgery.

RESULTS

Participant characteristics

The study included 320 participants with varying levels of surgical experience and proficiency. Surgical experience levels comprised pre-medical students interested in surgery ($n = 1$, 0.3%), medical students interested in surgery ($n = 96$, 30.0%), surgical trainees ($n = 98$, 30.6%), attending surgeons with < 5 years of experience ($n = 79$, 24.7%), attending surgeons with 5-10 years of experience ($n = 24$, 7.5%), and attending surgeons with > 10 years of experience ($n = 22$, 6.9%). Proficiency levels were categorized as NA ($n = 74$, 23.1%), beginner ($n = 68$, 21.3%), intermediate ($n = 132$, 41.3%), and advanced ($n = 46$, 14.4%). Participants reported primary specialties including general surgery ($n = 84$, 26.3%), neurosurgery ($n = 36$, 11.3%), orthopedics ($n = 12$, 3.8%), and others, with 36.6% ($n = 117$) classified as NA [Supplementary Table 1]. Regarding robotic platform use, 66 participants (20.6%) had used a robotic platform (primarily daVinci, $n = 46$), while 254 (79.4%) had not.

Mean CARS scale scores

The mean scores for the CARS scale variables ranged from 4.03 to 4.58 [Table 1]. Robotic stapler use had the highest mean score [mean (M) = 4.58, standard deviation (SD) = 0.983], followed by port placement ($M = 4.39$, $SD = 1.083$) and the intangibles (e.g., ergonomic advantage and performance; $M = 4.36$, $SD = 1.117$). Docking had the lowest mean score ($M = 4.03$, $SD = 1.452$). The robotic stapler use variable also had the lowest standard deviation, showing low variance within this parameter.

Table 1. Mean and standard deviation for each CARS scale variable

Variable	Mean	Standard deviation
Tissue dissection, blunt and sharp	4.27	0.932
Tissue handling and retraction	4.05	1.127
Robotic stapler use	4.58	0.983
Arm exchange/swapping	4.28	1.081
Endoscope/camera use	4.18	1.078
Intracorporeal suturing and tying	4.34	1.020
Wristed articulation	4.11	1.137
Port placement	4.39	1.083
Docking	4.03	1.452
The intangibles (i.e., ergonomic advantage and performance)	4.36	1.117

CARS: Competency-Based Assessment of Robotic Surgery Skills.

Differences across surgical experience levels

One-way ANOVA revealed significant differences in mean CARS scores across surgical experience levels for all scale variables ($P < 0.001$, except robotic stapler use, $P = 0.019$; Table 2). Surgical trainees and attending surgeons with > 10 years of experience generally reported higher scores. For example, tissue dissection scores were highest among surgical trainees ($M = 4.58$, $SD = 1.23$) and attending surgeons with > 10 years of experience ($M = 4.50$, $SD = 0.85$), compared to medical students ($M = 4.05$, $SD = 0.89$). Attending surgeons with < 5 years of experience consistently scored lower across variables (e.g., docking: $M = 2.48$, $SD = 0.98$).

Differences across proficiency levels

ANOVA results indicated significant differences in mean CARS scores across proficiency levels for all variables ($P \leq 0.001$; Table 3). Advanced proficiency participants reported the highest scores (e.g., intangibles: $M = 4.87$, $SD = 0.85$), while intermediate participants scored lowest across most variables (e.g., docking: $M = 3.12$, $SD = 1.30$). The NA group, possibly reflecting unclassified or self-reported non-proficient participants, scored consistently high (e.g., tissue dissection: $M = 5.00$).

Spearman's correlation analyses

Spearman's rank correlation analyses assessed associations between proficiency level and other variables [Table 4]. A moderate positive correlation was found between proficiency level and surgical experience ($\rho = 0.314$, $P < 0.001$), indicating that higher surgical experience was associated with greater proficiency. Proficiency level also showed moderate positive correlations with comfort using the CARS scale ($\rho = 0.337$, $P < 0.001$) and comfort using the CARS scale with feedback provided ($\rho = 0.329$, $P < 0.001$). The correlation between proficiency level and ever using a surgical robotic platform was weak and non-significant ($\rho = 0.107$, $P = 0.056$).

Regression analyses

Multivariate Linear Regression Analysis identified predictors of high CARS scale scores [Tables 5-9]. When examining surgical experience, significant predictors included port placement ($\beta = 0.35$, $P < 0.001$), endoscope/camera use ($\beta = -0.35$, $P = 0.001$), and docking ($\beta = -0.21$, $P = 0.018$), with the model explaining significant variance (e.g., docking: $F = 15.00$, $P < 0.001$). Robotic stapler use was a marginal predictor ($\beta = 0.13$, $P = 0.056$). For the robotic platform type, tissue handling and retraction was a significant predictor ($\beta = 0.29$, $P = 0.008$). For primary specialty, docking ($\beta = -0.42$, $P < 0.001$) and robotic stapler use ($\beta = 0.20$, $P = 0.004$) were significant predictors, with models explaining substantial variance (e.g., docking: $F = 47.40$, $P < 0.001$).

Table 2. ANOVA to compare the mean scores of the CARS scale across surgical experience

	Pre-medical student interested in surgery (n = 1)	Medical student interested in surgery (n = 96)	Surgical trainee (n = 98)	Attending surgeon with < 5 years of experience (n = 79)	Attending surgeon with 5-10 years of experience (n = 24)	Attending surgeon with > 10 years of experience (n = 22)	P value
Tissue dissection, blunt and sharp	5	4.05 ± 0.89	4.58 ± 1.23	4.06 ± 0.40	4.29 ± 0.46	4.50 ± 0.85	< 0.001
Tissue handling and retraction	5	3.99 ± 1.01	4.60 ± 1.27	3.35 ± 0.73	4 ± 0.83	4.36 ± 1	< 0.001
Robotic stapler use	5	4.18 ± 1.04	4.72 ± 1.18	4.80 ± 0.54	4.63 ± 0.49	4.82 ± 0.90	0.019
Arm exchange/swapping	5	4.05 ± 1.16	4.65 ± 1.30	3.97 ± 0.48	4.25 ± 0.79	4.73 ± 0.88	< 0.001
Endoscope/camera use	4	4.38 ± 0.92	4.62 ± 1.19	3.29 ± 0.66	4.17 ± 0.63	4.55 ± 0.96	< 0.001
Intracorporeal suturing and tying	3	4.23 ± 0.97	4.66 ± 1.26	3.97 ± 0.57	4.38 ± 0.71	4.77 ± 1.02	< 0.001
Wristed articulation	3	4.14 ± 0.99	4.53 ± 1.33	3.33 ± 0.77	4.50 ± 0.65	4.50 ± 1.05	< 0.001
Port placement	3	4.18 ± 1.07	4.74 ± 1.25	4.01 ± 0.49	4.38 ± 1.05	5.18 ± 1.05	< 0.001
Docking	4	4.33 ± 1.13	4.74 ± 1.22	2.48 ± 0.98	4.08 ± 1.21	5 ± 1.15	< 0.001
The intangibles (i.e., ergonomic advantage and performance)	5	4.30 ± 1.08	4.67 ± 1.34	3.92 ± 0.63	4.42 ± 0.83	4.68 ± 1.28	< 0.001

Parameters with a statistical significance level of $P < 0.05$ are highlighted in bold. ANOVA: Analysis of variance; CARS: Competency-Based Assessment of Robotic Surgery Skills.

Comparison of mean CARS scores across groups

Mean CARS scores varied by proficiency, surgical experience, and robotic platform use [Table 10]. Advanced proficiency participants had the highest mean score ($M = 4.04$, $SD = 0.34$, 95%CI [3.93, 4.14]) (CI = confidence interval), while beginners scored lowest ($M = 2.94$, $SD = 1.51$, 95%CI [2.57, 3.31]). Box plots confirmed significant differences across proficiency levels ($P < 0.001$ for beginner vs. advanced and intermediate vs. advanced; $P = 0.020$ for beginner vs. intermediate; Figure 1). Among surgical experience groups, attending surgeons with 5-10 years of experience ($M = 3.82$, $SD = 0.35$, 95%CI [3.67, 3.97]) and medical students ($M = 3.65$, $SD = 0.78$, 95%CI [3.49, 3.81]) scored higher than surgical trainees ($M = 2.55$, $SD = 1.78$, 95%CI [2.19, 2.91]). Significant differences were observed across surgical experience levels ($P < 0.001$; Figure 2). Participants who had used a robotic platform scored higher ($M = 3.69$, $SD = 0.08$, 95%CI [3.49, 3.88]) than those who had not ($M = 3.10$, $SD = 1.32$, 95%CI [2.93, 3.26]).

DISCUSSION

This study introduces the CARS scale, evaluating its potential to assess robotic surgical skills among 320 participants with varying experience and proficiency levels, and providing a comprehensive evaluation of global perceptions and the feasibility of using the CARS scale for video-based assessments of robotic surgical skills. By engaging a broad cohort of evaluators across varying levels of surgical training and specialties, we explored the comfort, perceived reliability,

Table 3. ANOVA to compare the mean scores of the CARS scale across proficiency levels

	NA (n = 74)	Beginner (n = 68)	Intermediate (n = 132)	Advanced (n = 46)	P value
Tissue dissection, blunt and sharp	5	4.16 ± 1.27	4.08 ± 0.53	4.41 ± 0.54	0.001
Tissue handling and retraction	5	4.12 ± 1.34	3.55 ± 0.83	4.50 ± 0.58	< 0.001
Robotic stapler use	5	4.22 ± 1.26	4.58 ± 0.72	4.76 ± 0.43	0.001
Arm exchange/swapping	5	4.21 ± 1.38	4.02 ± 0.79	4.52 ± 0.65	< 0.001
Endoscope/camera use	4	4.40 ± 1.18	3.66 ± 0.86	4.35 ± 0.73	< 0.001
Intracorporeal suturing and tying	3	4.25 ± 1.27	4.05 ± 0.73	4.72 ± 0.62	< 0.001
Wristed articulation	3	4.26 ± 1.27	3.58 ± 0.89	4.48 ± 0.75	< 0.001
Port placement	3	4.28 ± 1.24	4.01 ± 0.82	4.83 ± 0.95	< 0.001
Docking	4	4.34 ± 1.28	3.12 ± 1.30	4.78 ± 1.05	< 0.001
The intangibles (i.e., ergonomic advantage and performance)	5	4.24 ± 1.24	3.99 ± 0.78	4.87 ± 0.85	< 0.001

Parameters with a statistical significance level of $P < 0.05$ are highlighted in bold. ANOVA: Analysis of variance; CARS: Competency-Based Assessment of Robotic Surgery Skills.

Table 4. Spearman's correlation

	Surgical experience levels	Ever used a surgical robotic platform
	ρ 0.314	0.107
Proficiency level	$P < 0.001$	0.056
	N 320	320
	Comfortable using the CARS scale	Comfortable using the CARS scale if feedback provided
	ρ 0.337	0.329
	$P < 0.001$	< 0.001
	N 320	320

Parameters with a statistical significance level of $P < 0.05$ are highlighted in bold.

Table 5. Regression analysis to determine the predictors of high CARS scale scores with surgical experience level

	Coeff	Std. Err.	Beta	t Stat	P value
Tissue dissection, blunt and sharp	0.11	0.12	0.08	0.88	0.376
Tissue handling and retraction	-0.55	0.10	-0.05	-0.52	0.604
Robotic stapler use	0.16	0.08	0.13	1.19	0.056
Arm exchange/swapping	0.08	0.08	0.08	0.99	0.320
Endoscope/camera use	-0.38	0.11	-0.35	-3.33	0.001
Intracorporeal suturing and tying	0.03	0.11	0.02	0.26	0.788
Wristed articulation	0.06	0.09	0.06	0.65	0.511
Port placement	0.38	0.10	0.35	3.85	< 0.001
Docking	< 0.17	0.07	-0.21	-2.36	0.018
The intangibles (i.e., ergonomic advantage and performance)	-0.11	0.08	-0.10	-1.25	0.210

Parameters with a statistical significance level of $P < 0.05$ are highlighted in bold. CARS: Competency-Based Assessment of Robotic Surgery Skills.

Table 6. Regression analysis to determine the predictors of high CARS scale scores with surgical experience level

		Sum of squares	gl	Mean Square	F	P value
Robotic stapler use	Regression	17.33	1	17.33	12.89	< 0.001
	Residual	427.46	318	1.34		
	Total	444.79	319			
Endoscope/camera use	Regression	43.85	2	21.92	12.33	< 0.001
	Residual	400.94	317	1.26		
	Total	444.79	319			
Port placement	Regression	61.63	3	20.54		
	Residual	383.16	316	1.21	16.94	< 0.001
	Total	444.79	319			
Docking	Regression	71.178	4	17.79	15.00	< 0.001
	Residual	373.61	315	1.18		
	Total	444.79	319			

Parameters with a statistical significance level of $P < 0.05$ are highlighted in bold. CARS: Competency-Based Assessment of Robotic Surgery Skills.

Table 7. Regression analysis to determine the predictors of high CARS scale scores with robotic platform type used

	Coeff	Std. Err.	Beta	t Stat	P value
Tissue dissection, blunt and sharp	-0.01	0.08	-0.02	-0.21	0.830
Tissue handling and retraction	0.18	0.06	0.29	2.69	0.008
Robotic stapler use	-0.03	0.05	-0.05	-0.69	0.486
Arm exchange/swapping	-0.07	0.05	-0.11	-1.33	0.187
Endoscope/camera use	-0.04	0.07	-0.06	-0.58	0.561
Intracorporeal suturing and tying	-0.11	0.07	-0.17	-1.64	0.101
Wristed articulation	-0.07	0.06	-0.122	-1.24	0.215
Port placement	0.11	0.06	0.18	1.86	0.063
Docking	0.03	0.04	0.06	0.62	0.534
The intangibles (i.e., ergonomic advantage and performance)	0.02	0.05	0.04	0.45	0.653

CARS: Competency-Based Assessment of Robotic Surgery Skills.

Table 8. Regression analysis to determine the predictors of high CARS scale scores with primary specialty

	Coeff	Std. Err.	Beta	t Stat	P value
Tissue dissection, blunt and sharp	-0.46	0.68	-0.06	-0.67	0.503
Tissue handling and retraction	0.40	0.58	0.06	0.69	0.490
Robotic stapler use	1.37	0.47	0.20	2.93	0.004
Arm exchange/swapping	0.12	0.48	0.02	0.24	0.806
Endoscope/camera use	-1.02	0.63	-0.16	-1.61	0.107
Intracorporeal suturing and tying	0.14	0.61	0.02	0.22	0.820
Wristed articulation	0.15	0.52	0.02	0.28	0.774
Port placement	0.69	0.54	0.11	1.26	0.207
Docking	-1.93	0.41	-0.42	-4.71	< 0.001
The intangibles (i.e., ergonomic advantage and performance)	0.01	0.48	0.002	0.02	0.977

Parameter with a statistical significance level of $P < 0.05$ is highlighted in bold. CARS: Competency-Based Assessment of Robotic Surgery Skills.

Table 9. Regression analysis to determine the predictors of high CARS scale scores with primary specialty

		Sum of squares	gl	Mean Square	F	P value
Docking	Regression	1,790.88	1	1,790.88	47.40	< 0.001
	Residual	12,013.03	318	37.77		
	Total	13,803.92	319			
Robotic stapler use	Regression	2,494.77	2	1,247.38	34.96	< 0.001
	Residual	11,309.15	317	35.67		
	Total	13,803.92	319			

Parameters with a statistical significance level of $P < 0.05$ are highlighted in bold. CARS: Competency-Based Assessment of Robotic Surgery Skills.

Table 10. Comparison tables for mean CARS scores across different groups

	Total	Mean CARS score	Est. Desv.	95%CI
Proficiency levels				
Beginner	68	2.94	1.51	2.57-3.31
Intermediate	132	3.40	0.05	3.29-3.50
Advanced	46	4.04	0.34	3.93-4.14
Surgical experience				
Pre-medical student interested in surgery	1	3.82	0	0
Medical student interested in Surgery	96	3.65	0.78	3.49-3.81
Surgical trainee (fellow, resident, registrar, intern, postgraduate year...)	98	2.55	1.78	2.19-2.91
Consultant/attending level surgeon with < 5 years of experience	79	3.28	0.46	3.18-3.39
Consultant/attending level surgeon with 5-10 years of experience	24	3.82	0.35	3.67-3.97
Consultant/attending level surgeon with > 10 years of experience	22	3.38	1.40	2.76-4
Robotic platform type used				
daVinci	46	3.59	0.89	3.32-3.86
daVinci + others	3	4.05	0.28	3.36-4.75
Others	14	3.80	0.52	3.49-4.10
Ever used a robotic platform				
Yes	66	3.69	0.08	3.49-3.88
No	254	3.10	1.32	2.93-3.26

CARS: Competency-Based Assessment of Robotic Surgery Skills; CI: confidence interval.

and effectiveness of the CARS scale in modern competency-based surgical education. Key findings included mean CARS scores ranging from 4.03 to 4.58, with the robotic stapler task achieving the highest mean score (4.58) and the lowest standard deviation (0.983), as well as the differences across surgical experience and proficiency levels, with more experienced professionals receiving higher scores. Moreover, we conducted a Spearman's correlation analysis, obtaining a moderate positive correlation between proficiency level, surgical experience, and comfort. Finally, we evaluate mean CARS scores across groups where significant differences were observed across surgical experience levels. We consider that these observed findings further highlight the CARS scale's potential to ultimately optimize surgical outcomes, positioning it as a valuable tool compared to existing scales such as GEARS, RO-SCORE, R-OSATS, and ARCS.

The CARS scale offers key enhancements over existing robotic surgery assessment instruments, particularly in its domain specificity, broader applicability, and clinical relevance. The Global Evaluative Assessment of Robotic Skills (GEARS) is among the most widely cited tools, having demonstrated construct validity in

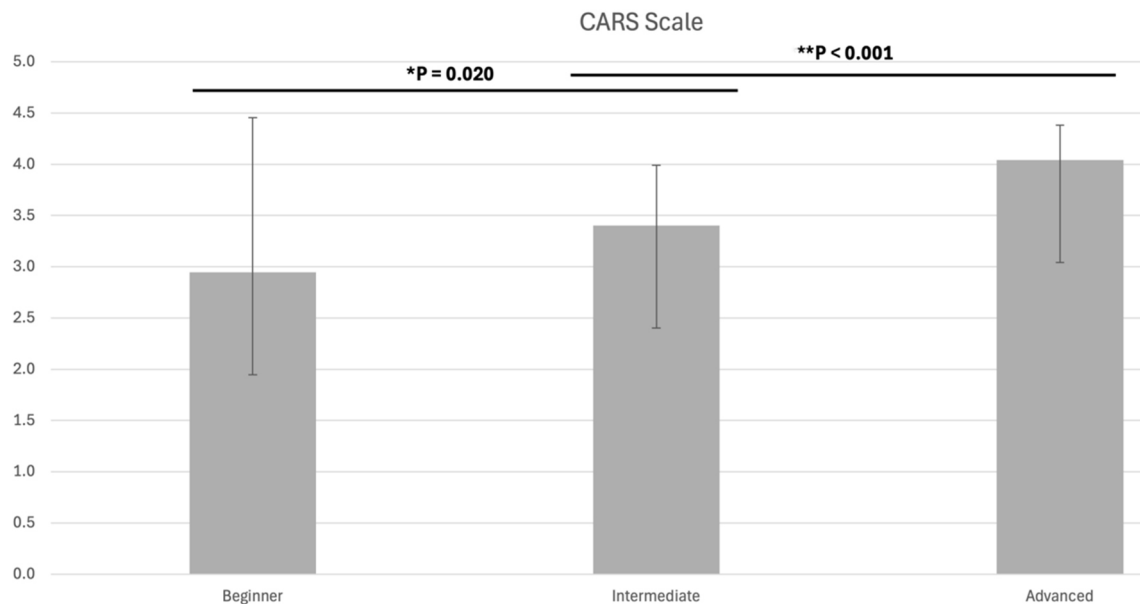


Figure 1. Box plot of mean CARS scores across proficiency levels (beginner, intermediate, advanced). Significant differences are indicated (* $P = 0.020$, ** $P < 0.001$). CARS: Competency-Based Assessment of Robotic Surgery Skills.

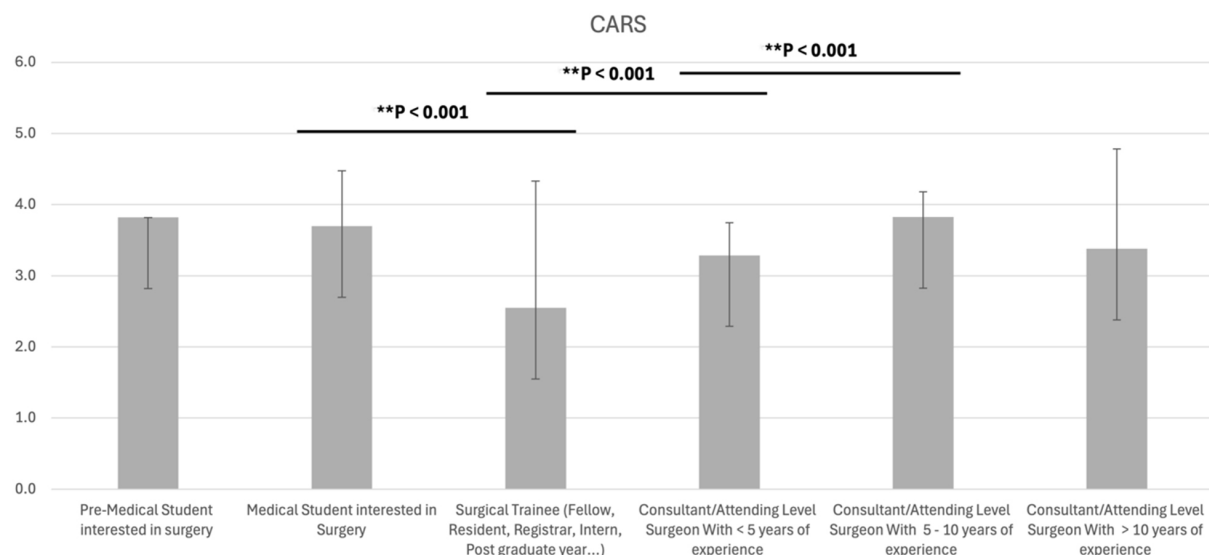


Figure 2. Box plot of mean CARS scores across surgical experience levels (pre-medical student, medical student, surgical trainee, attending surgeon with < 5 years, 5-10 years, and > 10 years of experience). Significant differences are indicated (** $P < 0.001$).

clinical settings^[12,17,18]. However, GEARS provides only generalized descriptors for skill categories, potentially limiting its sensitivity to procedure-specific technical nuances such as docking or intracorporeal suturing. In contrast, CARS incorporates discrete scoring for these individual domains (e.g., docking: $M = 4.03$; suturing: $M = 4.34$), thus enabling more precise benchmarking in competency-based curricula.

While GEARS has primarily been validated in urological and general surgery contexts, CARS has demonstrated utility across multiple specialties, including neurosurgery and general surgery, suggesting

broader translational potential. For instance, in this project, 11.3% of CARS participants were neurosurgeons, and regression analyses confirmed specialty-specific skill variation (e.g., docking $\beta = -0.42$, $P < 0.001$), which existing tools may not be calibrated to detect.

The Robotic Ottawa Surgical Competency Operating Room Evaluation (RO-SCORE), adapted from the O-SCORE^[19], evaluates independence rather than technical precision^[20,21]. Although autonomy is a valuable metric, it may fail to capture subtleties in technique or ergonomic competence. CARS addresses this limitation by differentiating across proficiency levels (e.g., advanced: $M = 4.04$ vs. beginner: $M = 2.94$) and identifying domain-specific deficits (e.g., wide variation in docking scores, $SD = 1.45$), reinforcing its role as a granular and formative evaluation tool.

Similarly, the Robotic Objective Structured Assessment of Technical Skills (R-OSATS), developed primarily for gynecologic training, has been validated only in dry-lab simulations^[14], limiting its external validity in real-world clinical settings. CARS, in contrast, was applied in active operative environments with a heterogeneous cohort and captured nuanced inter-specialty differences, enhancing its generalizability.

Lastly, the Assessment of Robotic Console Skills (ARCS), developed by Intuitive Surgical, has demonstrated construct validity in tissue models^[13], but its narrow focus on console maneuvers omits critical ergonomic and contextual elements of robotic performance. CARS uniquely includes a domain for ergonomic advantages ($M = 4.36$), a key yet underexplored component influencing fatigue, posture, and long-term surgical wellness. Furthermore, CARS' demonstrated sensitivity to training stage (e.g., surgical trainees: $M = 2.55$ vs. attending surgeons with 5-10 years: $M = 3.82$) underscores its potential as both an evaluative and educational scaffold in robotic surgery training programs.

Compared to CARS Phase 1, current findings reflect enhanced rubric precision, rater consistency, and expanded applicability across specialties and settings^[16]. Adding virtual simulation, asynchronous review, and scalable digital delivery in CARS 2.0 addresses prior limitations and opens avenues for global application in education environments constrained by geography or faculty availability.

Furthermore, this study highlights a broader philosophical shift within surgical education from time-based, subjective mentorship toward standardized, equitable assessments. CARS 2.0 was not merely a technical iteration but an embodiment of a new paradigm: that competency should be independent of location, and that structured feedback must not depend on proximity to academic centers. As Kallon (2024) emphasized in the context of global education, bridging the digital divide requires more than just access; it requires intentionally designed systems that empower marginalized learners to engage meaningfully with technology^[22]. This effort reimagines equitable surgical training in remote and resource-limited settings. CARS 2.0, centered on data collection, video evaluation, and feedback loop analysis, demonstrates how video-based competency scoring can enhance surgical education equity by delivering consistent, rubric-based feedback to trainees in both urban centers and underserved regions. It is an innovative assessment tool for scalable and inclusive training, not a replacement for traditional methods, but a complement to existing systems.

Implications for surgical training, limitations and future research aspects

The CARS scale's comprehensive evaluation of technical and intangible skills positions it as a promising tool for competency-based surgical training. Its sensitivity to experience and proficiency differences, along with its inclusion of specific metrics such as docking and suturing, addresses gaps in existing scales, making it well-suited for general surgery residency programs. The findings in this project indicate a moderate

correlation between proficiency and experience ($\rho = 0.314$), which underscores the need for structured training to bridge skill gaps, particularly for early-career surgeons who scored lower (e.g., attending surgeons with < 5 years: $M = 2.48$ for docking).

However, limitations exist. The NA proficiency group's high scores (e.g., tissue dissection: $M = 5.00$) suggest potential classification errors, which future studies should address. While CARS was applied in a diverse cohort, its validation in clinical settings and across a wider range of procedures is needed. Future research should also explore its predictive validity for surgical outcomes, such as complication rates, and compare it directly with GEARS, RO-SCORE, R-OSATS, and ARCS in the same cohort to establish relative strengths. Most importantly, our primary aim is to recognize that the true clinical relevance of CARS depends on its applicability beyond single specialties. Future research should therefore examine how surgeons from different disciplines, training levels, and degrees of familiarity with robotic platforms interpret and use the scale, and whether cross-specialty perspectives influence scoring. Such analyses will help determine whether CARS can serve not only as a specialty-specific tool but also as a standardized framework for evaluating and comparing robotic skill development across different surgical disciplines. To expand its utility and generalizability, future directions should include multi-institutional validation across diverse training settings, integration with simulation platforms and digital portfolios, AI-assisted video scoring to reduce rater burden, standardization in rural and global centers to advance digital health equity, longitudinal tracking of skill progression, and adaptation to other subspecialties to assess its cross-specialty scalability.

In conclusion, the CARS scale effectively discriminated levels of robotic surgical proficiency across multiple training stages, specialties, and experience groups. Robotic stapler use and port placement emerged as the strongest-performing domains, while docking and endoscope handling remained challenging, particularly for intermediate learners and early-career surgeons. Advanced participants consistently demonstrated superior performance, whereas prior exposure to robotics alone did not predict higher scores, underscoring the importance of structured assessment and targeted training. CARS scale offers a robust competency-based framework for evaluating robotic surgical skills, surpassing existing tools such as GEARS, RO-SCORE, R-OSATS, and ARCS in specificity and applicability. By addressing critical gaps in standardization, it supports evolving educational needs, particularly in minimally invasive gastrointestinal surgery. Incorporating CARS into surgical curricula could enhance training quality, promote equitable and reproducible assessment, ensure competency in robotic surgery, and ultimately contribute to improved patient outcomes.

DECLARATIONS

Acknowledgments

This study was conducted on behalf of TROGSS - The Robotic Global Surgical Society. The authors extend gratitude to the society and its members for their support and commitment to advancing robotic surgical education and research.

Authors' contributions

Conceptualization: Goyal A, Rivero-Moreno Y, Oviedo RJ

Methodology: Goyal A, Acosta S, Ibarra K, Macias CA

Software: Park J, Koutentakis M, Rivero-Moreno Y

Validation: Fuentes-Orozco C, Vargas VPS, Koutentakis M

Formal analysis: Goyal A, Rivero-Moreno Y, Garcia A, Meza KDM, Fuentes-Orozco C, Suárez-Carreón LO

Investigation: Babu A, Acosta S, Garcia A, Meza KDM, Fuentes-Orozco C, Rodriguez E

Resources: González-Ojeda A, Abou-Mrad A, Marano L

Data curation: Rivero-Moreno Y, Rodriguez E, Babu A

Writing - original draft preparation: Goyal A, Acosta S, Vargas VPS

Writing - review and editing: Ibarra K, Park J, Fuentes-Orozco C, Abou-Mrad A

Revision, editing, and rewriting: Koutentakis M, Marano L, Oviedo RJ, Goyal A

Visualization: Macias CA, Vargas VPS, Acosta S, Garcia A, Meza KDM, Fuentes-Orozco C, González-Ojeda A

Supervision: Suárez-Carreón LO, Marano L, Oviedo RJ

Project administration: Oviedo RJ, Abou-Mrad A, Goyal A

During the survey-based data collection, some members of On Behalf Of TROGSS, Collaborative Research Consortium provided their names while others provided only their E-mails [[Supplementary Materials](#)].

All authors have read and agreed to the published version of the manuscript.

Availability of data and materials

The anonymized surgical robotic video used for the assessment of the Competency Assessment in Robotic Surgery (CARS) scale in this study is available as [Supplementary Materials](#) accompanying this publication. All responses collected during the study are anonymous and are available from the corresponding author upon reasonable request. No personally identifiable information was collected or used.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

This study involved the anonymous analysis of a blinded, pre-recorded video for educational and research purposes. No personal identifiers or patient data were used or collected, and no intervention was performed. As such, ethical approval was not required in accordance with the institutional policies and the guidelines of the Declaration of Helsinki. All procedures complied with applicable ethical standards for research involving non-identifiable human data. All participants were required to provide electronic informed consent before participating in the study. Informed consent was implied by voluntary participation, with submission of de-identified responses constituting agreement to participate.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Liu Y, Feng Q, Xu J. Robotic surgery in gastrointestinal surgery: history, current status, and challenges. *Intell Surg.* 2024;7:101-4. [DOI](#)
2. Pugin F, Bucher P, Morel P. History of robotic surgery: from AESOP® and ZEUS® to da Vinci®. *J Visc Surg.* 2011;148:e3-8. [DOI](#) [PubMed](#)
3. Drake JM, Joy M, Goldenberg A, Kreindler D. Computer- and robot-assisted resection of thalamic astrocytomas in children. *Neurosurgery.* 1991;29:27-33. [DOI](#) [PubMed](#)
4. Kwok YS, Hou J, Jonckheere EA, Hayati S. A robot with improved absolute positioning accuracy for CT guided stereotactic brain surgery. *IEEE Trans Biomed Eng.* 1988;35:153-60. [DOI](#) [PubMed](#)
5. Morrell ALG, Morrell-Junior AC, Morrell AG, et al. The history of robotic surgery and its evolution: when illusion becomes reality.

- Rev Col Bras Cir.* 2021;48:e20202798. DOI PubMed PMC
6. Cornejo J, Cornejo J, Vargas M, et al. SY-MIS project: biomedical design of endo-robotic and laparoscopic training system for surgery on the earth and space. *Emerg Sci J.* 2024;8:372-93. DOI
 7. Picozzi P, Nocco U, Labate C, et al. Advances in robotic surgery: a review of new surgical platforms. *Electronics.* 2024;13:4675. DOI
 8. Müller DT, Ahn J, Brunner S, et al. Ergonomics in robot-assisted surgery in comparison to open or conventional laparoendoscopic surgery: a narrative review. *IJAWHS.* 2023;6:61-6. DOI
 9. Kankanamge D, Wijeweera C, Ong Z, et al. Artificial intelligence based assessment of minimally invasive surgical skills using standardised objective metrics - a narrative review. *Am J Surg.* 2025;241:116074. DOI PubMed
 10. Chang OH, King LP, Modest AM, Hur HC. Developing an objective structured assessment of technical skills for laparoscopic suturing and intracorporeal knot tying. *J Surg Educ.* 2016;73:258-63. DOI PubMed
 11. Niitsu H, Hirabayashi N, Yoshimitsu M, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today.* 2013;43:271-5. DOI PubMed PMC
 12. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol.* 2012;187:247-52. DOI PubMed
 13. Liu M, Purohit S, Mazanetz J, Allen W, Kreaden US, Curet M. Assessment of Robotic Console Skills (ARCS): construct validity of a novel global rating scale for technical skills in robotically assisted surgery. *Surg Endosc.* 2018;32:526-35. DOI PubMed
 14. Siddiqui NY, Galloway ML, Geller EJ, et al. Validity and reliability of the robotic objective structured assessment of technical skills. *Obstet Gynecol.* 2014;123:1193-9. DOI PubMed PMC
 15. Kutana S, Bitner DP, Addison P, Chung PJ, Talamini MA, Filicori F. Objective assessment of robotic surgical skills: review of literature and future directions. *Surg Endosc.* 2022;36:3698-707. DOI PubMed
 16. Elzein SM, Corzo MP, Tomey D, et al. Development and initial experience of a novel Competency-Based Assessment of Robotic Surgery Skills (CARS) scale for general surgery residents. *Global Surg Educ.* 2024;3:265. DOI
 17. Sánchez R, Rodríguez O, Rosciano J, et al. Robotic surgery training: construct validity of Global Evaluative Assessment of Robotic Skills (GEARS). *J Robot Surg.* 2016;10:227-31. DOI PubMed
 18. Aghazadeh MA, Jayaratna IS, Hung AJ, et al. External validation of Global Evaluative Assessment of Robotic Skills (GEARS). *Surg Endosc.* 2015;29:3261-6. DOI PubMed
 19. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med.* 2012;87:1401-7. DOI PubMed
 20. Gerull W, Zihni A, Awad M. Operative performance outcomes of a simulator-based robotic surgical skills curriculum. *Surg Endosc.* 2020;34:4543-8. DOI PubMed
 21. Calvo H, Kim MP, Chihara R, Chan EY. A systematic review of general surgery robotic training curriculums. *Heliyon.* 2023;9:e19260. DOI PubMed PMC
 22. Kallon A. Perspective chapter: the digital divide in a global emergency - how technology is shaping academic performance in diverse communities. In: Dias DDS, Candeias M, Meltzoff K, Editors. *Academic performance - students, teachers and institutions on the stage.* London: IntechOpen; 2024. DOI