Development and Validation of Task-Specific Metrics for the Assessment of Linear Stapler-Based Small Bowel Anastomosis

Ganesh Sankaranarayanan, PhD, Lisa M Parker, MD, Kimberly Jacinto, MD, Doga Demirel, PhD, Tansel Halic, PhD, Suvranu De, ScD, James W Fleshman, MD, FACS

INTRODUCTION: Task-specific metrics facilitate the assessment of surgeon performance. This 3-phased study was designed to (1) develop task-specific metrics for stapled small bowel anastomosis, (2) obtain expert consensus on the appropriateness of the developed metrics, and (3) establish its discriminant validity.

METHODS: In Phase I, a hierarchical task analysis was used to develop the metrics. In Phase II, a survey of expert colorectal surgeons established the importance of the developed metrics. In Phase III, to establish discriminant validity, surgical trainees and surgeons, divided into novice and experienced groups, constructed a side-to-side anastomosis on porcine small bowel using a linear cutting stapler. The participants' performances were videotaped and rated by 2 independent observers. Partial least squares regression was used to compute the weights for the task-specific metrics to obtain weighted total score.

RESULTS: In Phase II, a total of 45 colorectal surgeons were surveyed: 28 with more than 15 years, 13 with 5 to 15 years, and 4 with less than 5 years of experience. The consensus was obtained on all the task-specific metrics in the more experienced groups. In Phase III, 20 subjects participated equally in both groups. The experienced group performed better than the novice group regardless of the rating scale used: global rating scale ($p = 0.009$) and the task-specific metrics ($p = 0.012$). After partial least squares regression, the weighted task-specific metric score continued to show that the experienced group performed better ($p < 0.001$).

CONCLUSION: Task-specific metric items were developed based on expert consensus and showed good discriminant validity compared with a global rating scale between experienced and novice operators. These items can be used for evaluating technical skills in a stapled small bowel anastomosis model. (J Am Coll Surg 2022;235:881–893. © 2022 by the American College of Surgeons. Published by Wolters Kluwer Health, Inc. All rights reserved.)

Assessment of technical skills is critical in surgery and can be used for feedback for trainees or as part of high-stakes certification to enter surgical practice. Although assessment could be performed by watching live procedures in the operating room[1] or through validated simulation drills,[2–6] video-based assessment (VBA) has been the most used in

CME questions for this article available at http://jacscme.facs.org

Disclosure Information: Authors have nothing to disclose. Timothy J Eberlein, Editor-in-Chief, has nothing to disclose. Ronald J Weigel, CME Editor, has nothing to disclose.

Disclosures outside the scope of this work: Dr Demirel is a paid consultant to AVRA Medical.

Support: Drs Sankaranarayanan, Demirel, Halic, Fleshman and De are supported by National Institutes of Health (NIH)/National Institute of Biomedical Imaging and Bioengineering Grant R01 EB025241.

Presented virtually at the American College of Surgeons 107th Annual Clinical Congress, October 2021 and at the American Society of Colon & Rectal Surgeons Annual Scientific Meeting, April 2021.

Received March 17, 2022; Revised June 15, 2022; Accepted July 28, 2022.

From the Department of Surgery, Baylor University Medical Center, Dallas, TX (Sankaranarayanan, Parker, Jacinto, Fleshman); Department of Computer Science, Florida Polytechnic University, Lakeland, FL (Demirel); Department of Computer Science, University of Central Arkansas, Conway, AR (Halic); Department of Mechanical, Aerospace and Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY (De).

Present address: Department of Surgery, UT Southwestern Medical Center, Dallas, TX (Sankaranarayanan).

Correspondence address: Ganesh Sankaranarayanan, PhD, Division of Bariatrics and Foregut Surgery, Department of Surgery, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, TX 75390-9169. email: ganesh.sankaranarayanan@UTSouthwestern.edu

Supplemental digital content is available for this article.

https://doi.org/10.1097/XCS.0000000000000389
ISSN 1879-1190/22

Abbreviations and Acronyms

COSATS	=	Colorectal Objective Structured Assessment of Technical Skill
HTA	=	hierarchical task analysis
ICC	=	intraclass correlation coefficient
PLS	=	partial least squares
VBA	=	video-based assessment
VCOST	=	Virtual Colorectal Surgery Trainer

practice.[7–9] VBA of technical skills has also been shown to predict outcomes and rate of complication.[10,11] Recently, the American Board of Surgery (ABS) has announced that they will be piloting a VBA assessment of technical skills as part of the board certification process.[12]

There are 2 major types of technical skills assessment: summative for high stakes and formative for providing feedback to trainees.[13] The global assessment of technical skills, such as the Objective Structured Assessment of Technical Skills (OSATS), assesses performance in 6 different domains and is used for the summative evaluation. For formative assessment, a more detailed assessment of each procedure step needs to be performed to provide feedback. We have previously used the hierarchical task analysis (HTA) to break down surgical procedures into detailed steps and developed metrics for assessment, demonstrating its validity (eg for cricothyrotomy[14] and double-layered hand-sewn anastomosis of the small bowel[15]).

Candidates for specialization in colon and rectal surgery train an additional year beyond general surgery training in a colorectal residency monitored by the Residency Review Committee for Colon and Rectal Surgery of the Accreditation Council for Graduate Medical Education. They subsequently undergo both a qualifying written and a certifying oral exam administered by the American Board of Colon and Rectal Surgery. In-training operative skills performance evaluation is based on subjective feedback by the faculties in the training program. Currently, the technical skills of trainees are not assessed objectively. Studies comparing these subjective evaluations with more objective assessments, such as task-specific metrics, have found they do not evaluate a residents' technical performance as well as the objective evaluations.[16,17] The operative assessment committee of the American Society of Colon and Rectal Surgeons (ASCRS) developed the Colorectal Objective Structured Assessment of Technical Skill (COSATS) to address the question of assessing technical competency of the colorectal surgery trainees. Their study combined both a global rating scale and a task-specific checklist for select tasks considered integral to the practice of an independent colorectal surgeon. They found that both scales could reliably differentiate colorectal surgery residents from general surgery residents.[18] Despite their merits, skill assessment tools require considerable cost, including examiner and staff time, labor, and materials to set up for each participant. Additionally, a trained proctor must administer and score the performance, which adds another barrier to participation.

To overcome the practical challenges of conducting a technical assessment exam, we are developing the Virtual Colorectal Surgery Trainer (VCOST), a virtual reality–based open colorectal surgery simulator with open surgical tasks of COSATS. We had previously developed and validated metrics for assessing skills in double-layered hand-sewn small bowel anastomosis.[19] The present work aims to create and validate metrics for the assessment of skills in anastomosis of small bowel using a linear stapler.

METHODS

Development of task-specific metrics for bowel anastomosis using a linear stapler

In Phase I of this IRB-approved study, a detailed HTA was constructed for the functional end-to-end small bowel anastomosis using a linear stapler by conducting interviews with expert colorectal surgeons at the Baylor University Medical Center (BUMC). This information was supplemented with procedural details available in surgery textbooks and by watching workshop videos. This procedure's major tasks and subtasks were identified and refined by

an expert colorectal surgeon (JWF). A 5-point scale was applied to each task-specific metric, using a Likert scale, with 5 being wholly correct and 1 being completely incorrect. Any intermediate points between 5 and 1 points were regarded as suboptimal performances.

In Phase II of the study, consensus on the importance of the developed metrics was obtained from practicing colorectal surgeons (less than 5 years, 5 to 15 years, more than 15 years of experience after completion of their training) using an online survey administered using the SurveyMonkey platform. Weighted averages of the ratings of each metric item on a 5-point Likert scale, with 1 being the least important and 5 the most important, were used to assess the importance of the developed metrics.

Validation of the metrics

In Phase III of the study, an IRB-approved validation study was conducted at BUMC in the Baylor Operative Skills Simulation (BOSS) Laboratory to assess the usefulness of the developed task-specific metrics in evaluating the difference in operative performance (discriminant validity) between experienced and novice operators. The participants were asked to perform a side-to-side anastomosis using a linear stapler (GIA Auto Suture Stapler, Coviden Inc.) on porcine small bowel with mesentery. The frozen porcine small bowel (sourced from Animal Technologies Inc.) was thawed and cut into 20-cm-long pieces. The small bowel segments were then further divided with a linear stapler in the center, resulting in 2 10-cm-long pieces of the bowel, with adjacent stapled ends in the center and open lumen at the segment's outer ends mesentery of the 20-cm segment intact.

The surgical participants were divided into a novice group (general surgery PGY 1–3) and an experienced group (PGY 4-5, colorectal surgery residents and attendings). At BUMC, the residents start their colorectal rotation at the end of PGY3, and hence the residents in the first 3 years of their training were assigned to the novice group. Some of them may be exposed to colorectal procedures as part of their trauma surgery rotations, but we do not precisely control for this exposure in our study. The anastomosis was constructed between the 2 stapled ends in the middle of the bowel segment to create a functional end-to-end/side-to-side anastomosis using the 60-mm linear cutter stapler.

The participants were deidentified, and their performance was videotaped using a digital camera (Sony Inc. HD Camera with 1080P resolution). Additionally, the positions of the surgeon's hands were tracked using magnetic tracking sensors (3D Guidance Ascension TrakStar). The hand motion data were not used in this analysis but will be used to analyze the workspace to develop a custom haptic device for our simulator. After the completion of the anastomosis, the small bowel specimen was inspected for tissue damage, the quality of the staple line, and the symmetrical placement of Lembert sutures used to invert the staple line.

A water pressure test was conducted on each specimen to assess anastomosis leak. A Kelly clamp was used to close 1 end of the bowel segment. A Tumi syringe containing 50cc of water was inserted into the opposite open end, and a second clamp was placed across the rest of the bowel opening on the same side. The water was inserted into the bowel to distend the bowel and perform the leak test. This allowed us to test for bowel patency across the anastomosis and document any visible leaks. Gentle manual pressure was also applied to assure the staple line was adequately distended with water. We did not conduct a burst pressure. Two qualified raters (L.M.P, K.J.) assessed the performance from the video recordings using the developed task-specific metrics and the global rating scale of performance shown in Table 1, which was previously developed and validated for the COSATS.[20,21]

Statistical analysis

In Phase II, consensus development of task-specific metrics, descriptive statistics were used to analyze the data from the survey, and a weighted average of importance scores from the survey was calculated for all the metric items. The responses from colorectal surgeons within the 3 experience levels (less than 5 years, 5 to 15 years, more than 15 years) were analyzed using the nonparametric Kruskal-Wallis test. Post hoc analysis of significant results was performed using the Wilcoxon rank-sum test. To control the false discovery rate owing to multiple comparisons, Benjamini and Hochberg correction[22] was used to calculate the adjusted p values with significance set at 0.05.

In the Phase III validation study, intraclass correlation coefficient (ICC) was computed for absolute agreement to assess the interrater reliability of the 2 raters grading the operative videos. Values between 0.75 and 0.9 were considered good and anything greater than that as excellent reliability.[23] The ICC was calculated after the raters graded the first 5 videos. Discrepancies were resolved, and the process was repeated for the second set of 5 videos. On completion of rating all 20 of the videos, we averaged the ratings from the 2 raters to calculate the global and task-specific scores.

The total global score for each video was computed by summing all the metric items in Table 1. The total

Table 1. Global Assessment Metrics Used for Evaluating Performance for Side-by-Side Small Bowel Anastomosis Using a Linear Stapler.

No.	Domain of surgical performance	Rating (5-point Likert scale)				
		1	2	3	4	5
1	Respect for tissue	Frequently used unnecessary force on tissue or caused damage	—	Careful handling of tissue but occasionally caused inadvertent damage though more than needed force	—	Consistent handling of tissue, minimizes damage through appropriate use of instruments and appropriate force
2	Time and motion	Many unnecessary moves	—	Efficient time/motion but some unnecessary moves	—	Clear economy of movement and maximum efficiency
3	Instrument handling	Repeatedly made tentative or awkward moves with instruments	—	Competent use of instruments but occasionally appeared stiff or awkward	—	Fluid moves with instruments and no awkwardness
4	Flow of operation	Frequently stopped procedure and seemed unsure of next move	—	Demonstrated some forward planning with reasonable progression of procedures	—	Obviously planned course of procedure with effortless flow from one move to the next
5	Knowledge of instruments	Deficient knowledge or not familiar with instruments	—	Knew some of the instrument required for the procedure	—	Demonstrated familiarity with all the instruments for the procedure
6	Knowledge of specific procedure	Deficient knowledge	—	Knew all important steps of procedure	—	Demonstrated familiarity with all aspects of procedure
7	Overall performance	Very poor	—	Competent	—	Expert level

task-specific metric score for each video was calculated using 2 methods. The first method calculated a total task-specific metric score by adding all the individual metrics. In the second method, a weighted sum was computed using the weights for individual task-specific metric items calculated using partial least squares (PLS) regression[24] for predicting the total global metrics score. PLS regression, which is a widely used statistical method in chemometrics, is used for predicting a set of dependent variables from a larger set of predictor variables using a latent variable approach to model the covariance structure between the 2 variables.[25,26] We have used the PLS regression technique successfully in developing assessment tools for endotracheal intubation and hand-sewn bowel anastomosis.[19,27] Spearman Rank Correlation test was used to compute the correlation between the total Global and Checklist scores. We report both the correlation coefficient (r_s) and the coefficient of determination (R^2) computed by fitting a linear regression model.

For total Global and Checklist scores, the normality of data was checked using the Shapiro-Wilk test.[28] If the data were normal, we used the *t*-test to compute the difference between the groups; otherwise, we used the nonparametric Mann-Whitney-Wilcoxon test. The statistical analyses were performed using the R statistical package (version 4.0.2).[29] The statistical significance level was set at 0.05.

RESULTS

Phase 1: HTA results

The small bowel stapled anastomosis has the following 6 significant tasks (Fig. 1):

1, **Transecting the intestine:** The intestine is transected using a linear stapler after opening a mesenteric window on the mesenteric side of the bowel at the point of chosen transaction to allow the linear stapler to be placed across the intestine in a mesenteric to anti-mesenteric direction. The stapler should be angled towards the vascularized intestine that will remain as part of the anastomosis to prevent ischemia of the antimesenteric corner of the staple line. This allows a wedge of the mesentery to be resected with the part of the bowel to be removed. The process is repeated at the other side of the planned resection segment angling the staple line towards vascularized tissue.

2. **Securing the enterotomy:** Amputating the corner of the staple line on the antimesenteric side of the transverse staple line requires the corner or apex of the staple line to be grasped while the enterotomy

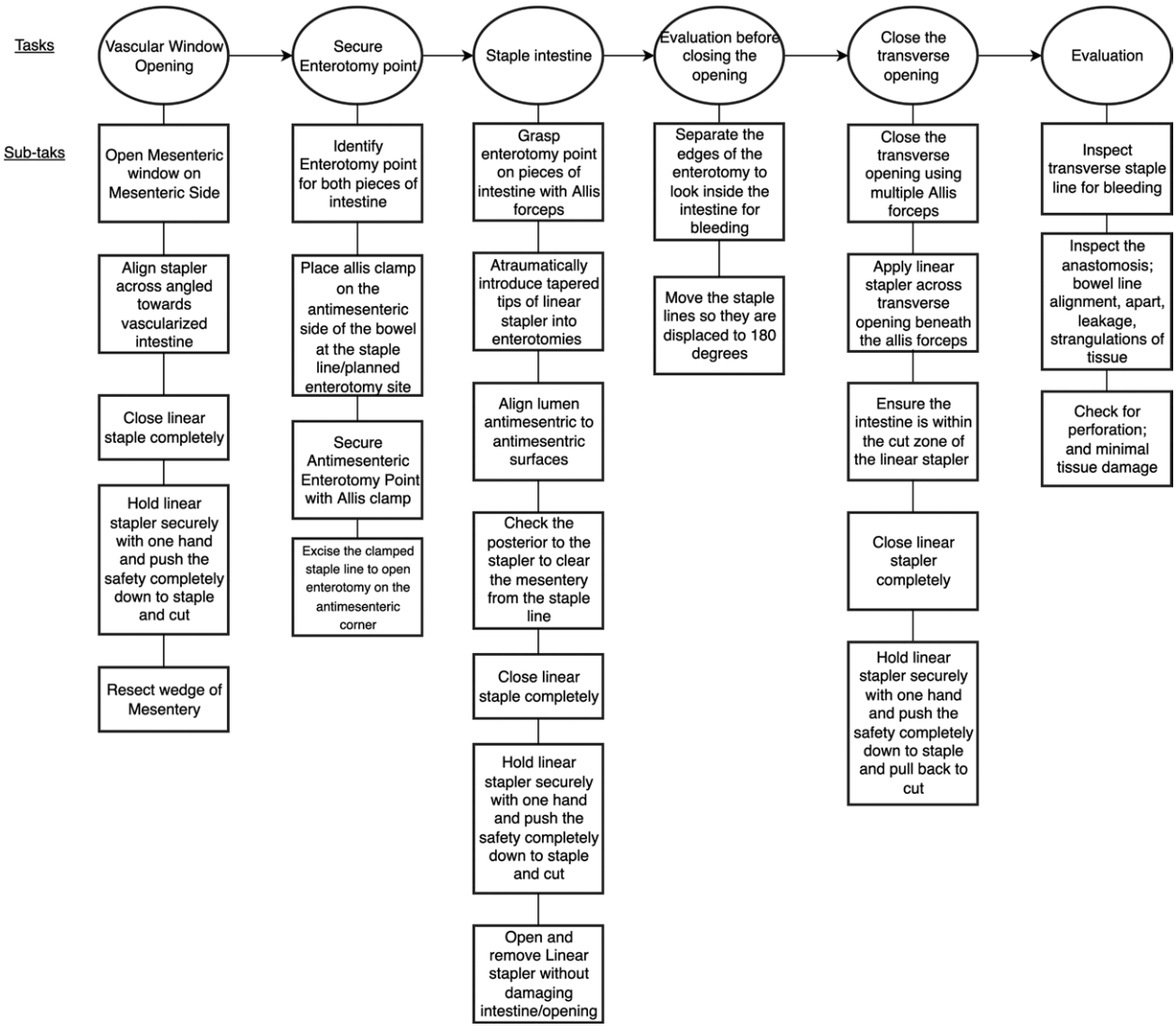

Figure 1. Hierarchical task analysis (HTA) of the small bowel anastomosis using a linear stapler.

is created and the staple line divided approximately 1 cm proximal to the apex. Securing the enterotomy using an allis clamp on the anti-mesenteric corner of the enterotomy allows both portions of the intestine to be aligned to place each portion of the linear stapler inside the two limbs of the bowel to create a side-to-side anastomosis.

3. **Stapling the intestine:** While the enterotomy points are grasped with allis forceps, tapered tips of the linear stapler is atraumatically placed into the enterotomy of each bowel piece, and both sides of the intestine are aligned at their anti-mesenteric surfaces. Careful inspection of the posterior aspect of the side-to-side anastomosis is needed to ensure that the mesentery layer is free of the staple line. Once the mesentery is free, the 2 arms of the stapler are rejoined. Stapler lengths such as 80 mm or 60 mm with 3.8-mm staple height are usually used for small bowel operation.[30] The safety is released on the handle to allow the cutting blade to be pushed towards the apex of the anastomosis, which also fires the staples as it cuts. The blade handle is pulled back to the starting position to open and remove the stapler. Handling and removing the stapler from the intestine should be done in such a way as to ensure no damage to the intestine or the opening.
4. **Evaluation before closing the opening:** The inside of the intestine needs to be examined through the transverse opening of the anastomosis to look for and control bleeding along the staple line. After this evaluation, the staple lines need to be displaced 180 degrees.

5. **Closing the transverse opening:** Several Allis forceps are used to securely reapproximate the cut edges of the common enterotomy with the 2 ends of the staple line placed as far apart as the enterotomy will allow. The linear stapler is placed across the transverse opening below the allis forceps on the outside of the bowel and fired. A Lembert suture is placed between the 2 limbs of the bowel at the apex of the staple line to protect the single staple holding the apex of the anastomosis together. The transverse opening staple line is over-sewn and buried with the equidistant placement of Lembert seromuscular sutures using absorbable sutures.
6. **Final evaluation:** Finally, the anastomosis requires careful examination for completeness, patency, mesenteric bleeding, tissue damage, and alignment of the staple line to avoid twisting or lumen compromise.

Phase II: Expert consensus survey results

A total of 45 colorectal surgeons participated in this survey, of whom 62.2 % (n = 28) had more than 15 years of experience, 28.89% (n = 13) had 5 to 15 years of experience, and 8.89% (n = 4) had less than 5 years of experience. For the question on adequate time to perform the anastomosis, the responses ranged from 1 to 60 minutes with an average of 14 minutes.

The agreed-on task-specific metrics for assessment of side-by-side stapled anastomosis are shown in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A135. The metrics are categorized into 2 major procedural categories: (1) Bowel transection/vascular window opening and (2) side-by-side anastomosis/closure of transverse opening, with (3) general metrics separately as a third category. The radar charts of the weighted averages of each of the 3 categories of the metrics grouped based on the 3 experience levels are shown in Figures 2a through 2c.

For the vascular window opening metrics (Fig 2a), the less than 5 years group rated all 7 metrics very important with weighted averages ranging from 4.67 to 5. The 5 to 15 years group weighted average ranged from 3.25 to 4.62, with the least for M5 (hold linear stapler securely) and the highest for M2 (intestine transection using linear stapler). The more than 15 years groups weighted averages ranged from 3.95 to 4.6 with the least for M2 (intestine transection using linear stapler) and the highest for M4 (close the linear stapler), The Kruskal-Wallis test showed no significant differences in the 3 groups' rating of importance.

For the side-by-side anastomosis metrics (Fig. 2b), the less than 5 years rated all the metrics high with the weighted average scores ranging from 4.33 to 5. The 5 to 15 years group weighted average ranged from 2.87 to 5, with the least for M27 (suture handling during oversewing of transverse staple line) and the highest for M25 (check transverse staple line). For the more than 15 years group, the weighted average ranged from 3.26 to 4.73 with the least for M28 (check the lumen), and the highest for M9 (align the intestine). The Kruskal-Wallis test showed a significant difference between the groups in their ratings for 2 metric items, M18 (p = 0.027) and M27 (p = 0.047). Post hoc analysis showed that for the metric item M18 (distract the staple lines so they are displaced to 180 degrees), there was a significant difference in rating between the less than 5 years and the 5 to 15 years group (p = 0.03) and no difference between the 5 to 15 years and the more than 15 years group (p = 0.13). For the metric item M27, no significant differences were found in post hoc comparisons.

For the general metrics (Fig. 2c), the less than 5 years group rated all the metric items high, with the weighted average scores ranging from 4.33 to 5. For the 5 to 15 years group, the weighted average scores ranged from 2.5 to 4.37, with the lowest for M31 (linear stapler size) and the highest for both M36 (knowledge of instruments and

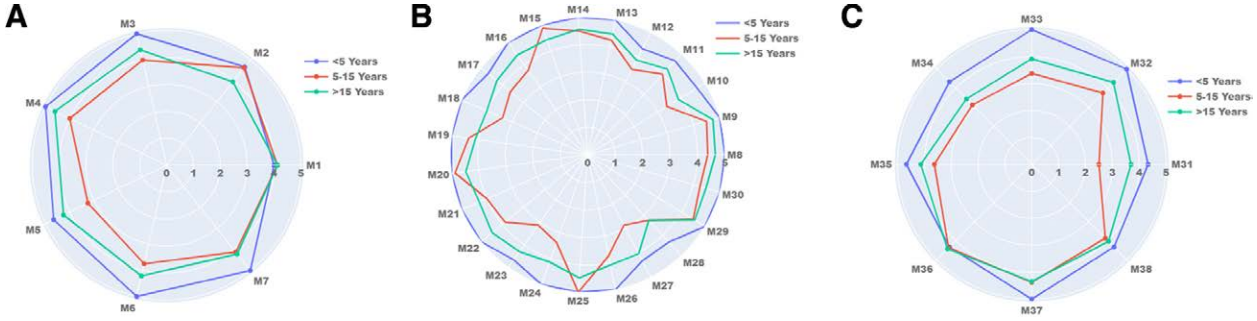

Figure 2. (A) Plot of weighted average scores for the vascular window opening metrics for the 3 groups. (B) Plot of weighted average scores for the side-by-side stapled anastomosis metrics for the 3 groups. (C) Plot of weighted average scores for the general metric for the 3 groups.

procedure) and M37 (tasks completion). For the less than 15 years group, the weighted averages ranged from 3.43 to 4.43, with the lowest for M34 (type of suture) and the highest for M36 (knowledge of instruments and procedure). The Kruskal-Wallis test showed no significant differences in the 3 groups' rating of importance.

Phase III: Metrics validation study results

A total of 20 participants, equally distributed between experienced and novice, participated in this study. The experienced group consisted of 4 attendings, 1 fellow, 2 PGY-5, and 3 PGY-4 residents, and the novice group consisted of 5 PGY-3, 2 PGY-2, and 3 PGY-1 residents. Because the recorded video for 1 subject from the experienced group was corrupted and unrecoverable, we used only 19 subjects for the assessment and rating by a third party using global and task-specific metrics. The leak test was performed for all 20 subjects. The experimental setup showing the porcine small bowel model and a subject performing the side-by-side anastomosis is shown in Figures 3a and 3b.

Reliability analysis

The ICC analysis for the ratings by our 2 raters showed that both global (ICC = 0.77, 95% CI 0.67–0.84, $p < 0.001$) and task-specific (ICC = 0.82, 95% CI 0.76–0.86, $p < 0.001$) metrics had good reliability between the raters.

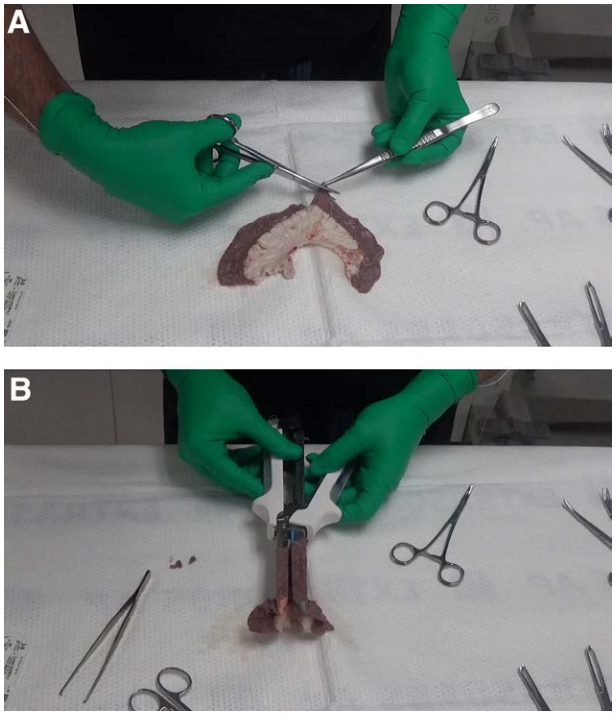

Figure 3. (A) Porcine small bowel model with excision being made to open the enterotomy point. (B) Side-by-side anastomosis performed using a linear stapler.

Global metrics

The total global score computed from the average of the 2 raters is shown in Figure 4a. The experienced group performed significantly better than the novice group (median score = 29.75 vs 19, W = 77, p = 0.009).

Task-specific metrics

We used only a subset of the developed metrics for the task-specific assessment. We used the explanted porcine intestine specimens, which were set up with 2 stapled ends aligned and placed for performing the anastomosis. Moreover, because there is no bleeding in the explanted porcine small bowel, the metrics associated with bleeding control were excluded. We also combined certain closely related metric items for assessment because it was easier to rate those specific steps accurately and achieve higher interrater reliability. Additionally, after observing the performance from the recorded videos, we added an additional metric item that assessed the adherence to the prescribed order of the task execution. The details of the final task-specific metric items and their order are shown in Table 2. The total task-specific score computed from the average of the 2 raters is shown in Figure 4b. The experienced group performed better than the novice group (median score = 68.5 vs 53.5, W = 76, p = 0.012). Figure 5a shows the correlation between the global and task-specific scores. The Spearman correlation coefficient showed moderate correlation with $r_s = 0.74$, $p = 0.002$ and $R^2 = 0.65$.

PLS regression

Table 2 shows the weights computed using PLS regression for the task-specific metrics. The metric item M11-12 corresponding to grasp and excise the enterotomy point on pieces of intestine had the highest weight of 1.0228. Three metric items, M24 (push the tab of the stapler completely down to staple and cut), M27 (suture handling during oversewing of transverse staple line), and M35 (overall suture handling), had negative weights. A new task-specific score was then calculated using the computed weights by adding each metric item multiplied by their weights. Figure 5b shows the significant improvement in correlation after the PLS regression with $r_s = 0.848$, $p < 0.001$, $R^2 = 0.84$. The task-specific metric computed using the PLS also showed that the experienced group performed

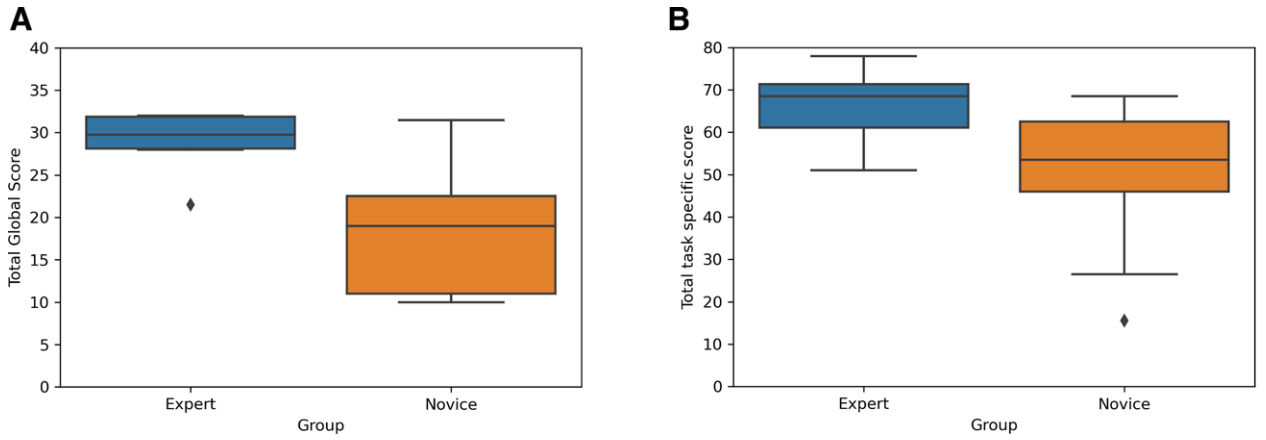

Figure 4. (A) Total global score for both the groups. (B) Total task-specific score for both the groups.

significantly better than the novice group (median score = 31.74 vs 21.81, W = 87, $p < 0.001$).

Task-specific metric reduction

Although we had 16 metric items for the performance assessment using our task-specific metrics, our interest is to find a minimal set that is sufficient to discriminate the performance between the 2 groups. To do this, we performed PLS regression by leaving out 1 metric item at a time to compute a new total score and repeated it for all the 16 metric items. With the newly computed total scores, the Mann-Whitney-Wilcoxon test was then used to compare the performance between the groups, as shown below in Table 3. We also report the R^2 value from the correlation between the new task-specific and global metric scores. Our analysis shows that the metric item M11-12 had the most impact in predicting the performance, and removing it negatively impacted the correlation with the global score. A Mann-Whitney-Wilcoxon test using the metric item M11-12 showed that the experienced group performed better than

Table 2. Weights Computed Using the Partial Least Squares Regression for Each of the Task-Specific Metric Items.

Task-specific metric item	Metric description	Weight computed using the PLS regression
M8	Identify and secure the enterotomy point for introducing the stapler tips for both pieces of intestine	0.6768
M11-12	Grasp and excise enterotomy point on pieces of intestine	1.454
M13	Introduce linear stapler into enterotomies	1.0228
M14-15	Alignment of the lumen as stapler is closed and check the hidden surface to clear mesentery from the staple line	0.6478
M4-5	Hold and close linear stapler	0.6895
M6	Push the tab completely down to staple and cut	0.2045
M16	Open and remove linear stapler	0.6436
M18	Distract and close the transverse opening	0.0361
M19-21	Linear stapler applied across the transverse opening beneath the Allis forceps to capture the intestine in the transverse staple line and within the cut zone of the linear stapler	0.323
M24	Push the tab completely down to staple and cut	−0.1233
M26	Place stitch at the apex of GIA staple line between limbs	0.6459
M27	Suture handling during oversewing of transverse staple line	−0.3987
M35	Overall suture handling	−0.0319
M37	Task completion	0.05738
M38	Completion time	0.6973
M39	Task execution order	0.4127

PLS, partial least squares.

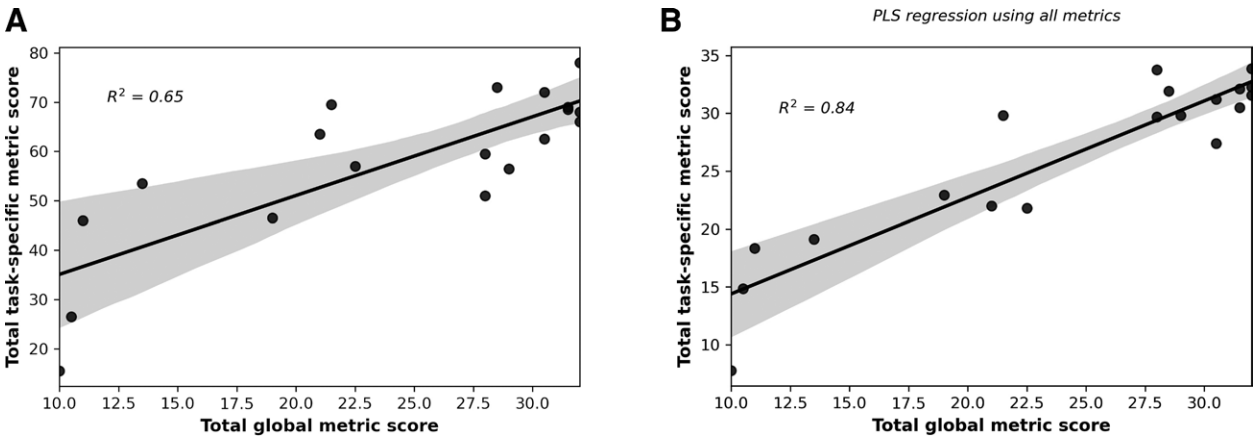

Figure 5. (A) Correlation between the total global score and total task specific metric score and a linear regression model fit with 95% confidence band. (B) Correlation between the total global score and total task-specific metric score computed using the partial least squares weight, and a linear regression model fit with 95% CI.

novices (median score = 5.0 vs. 1.5, W = 81.5, p=0.002). Correlation analysis show that it has a moderate correlation with the Global metric score (r_s = 0.728, p < 0.001, R^2 = 0.725).

Leak test results

A thorough inspection of the completed stapled small bowel specimen was conducted to check for tissue damage, quality of the sutures, spacing of the staple line, and whether any part of the mesentery was sutured. Figure 6 shows the specimen inflated with water to check for visible leaks. Specimens with positive leaks were recorded with a score of 1, and no visible leaks were given a score of 0. We observed anastomosis leak in 4 experienced and 5 novice specimens during the leak test. The chi-square test of independence showed no significant association between groups and the leak test outcome (chi-square = 0.202, p = 0.65).

DISCUSSION

Our study has shown that a global rating score or a task-specific metric rating can distinguish between experienced and novice operators in a simple task such as creating a side-to-side/functional end-to-end anastomosis

Table 3. Performance of Task-Specific Metrics Computed Using Partial Least Squares Regression with 1 Metric Item Removed at a Time.

Metric item removed	R^2 (global vs task-specific score)	Expert vs novice Mann-Whitney-Wilcoxon test (p value)
M8	0.846	0.007
M11-12	0.606	0.019
M13	0.836	<0.001
M14-15	0.826	<0.001
M4-5	0.837	<0.001
M6	0.844	<0.001
M16	0.844	<0.001
M18	0.84	<0.001
M19-21	0.833	<0.001
M24	0.833	<0.001
M26	0.845	<0.001
M27	0.83	<0.001
M35	0.844	<0.001
M37	0.841	<0.001
M38	0.838	<0.001
M39	0.838	<0.001

Figure 6. Leak test showing our set up with Kelly clamp to secure the openings and catheter through which 50cc of water is inserted.

in the porcine small intestine. The ability to determine which task-specific metrics truly discriminate between capable and noncapable individuals is also possible. This study also confirms that using a trained observer to rate the performance of a task on a video of a subject performing the task can determine the experience of an operator in the specific task of interest. Therefore, it is possible to create tasks in a simulation model that can discriminate and inform operator performance in a training program setting.

Surgical stapling devices such as the transverse staple line (TA), linear cutting stapler (GIA), and circular staplers (EEA) are routinely used to perform a portion of an anastomosis in colorectal surgery. The GIA linear stapler used in our study places 2 double staggered rows of titanium staples and divides the tissue in between. The stapler comes in 3 sizes of 60, 80, and 100 mm, with staple sizes of 2.5, 3.8, and 4.8 mm. In our study, the GIA stapler with 80-mm length and 3.8-mm staple size was used for performing the side-to-side anastomosis of the porcine small bowel. Colorectal anastomosis using a stapler is preferred for its shorter operating times and a lower chance of leakage. However, there is no evidence that it is better than the hand-sewn technique for preventing leaks.[31–34] The use of stapling devices is not without risk. In a study on reported fatalities and adverse events from the Manufacturer and User Facility Device Experience (MAUDE) database maintained by the FDA between 1992 and 2001, a total of 112 deaths, 180 injuries, and 22,804 malfunctions were recorded.[35] The gastrointestinal anastomotic site was responsible for 73 (65.2%) of 112 deaths. The complications reported include bleeding, suture line separation or leak, stapler not firing properly, stapler sticking or difficult to remove from the tissue, infection, and prolonged procedures owing to difficult anastomosis. Another review on complication using stapling devices has reported anastomotic leak, stricture, and bleeding as major complications to be anticipated and the rate of anastomotic stricture to be 4-fold higher than anastomosis with a hand-sewn technique.[36,37]

To reduce complication, it is critical that proper training in handling the stapler device and its application is needed. Moreover, both summative and formative assessment tools are needed to assess proficiency and provide task-specific feedback for improvement. The metrics developed to assess performance in our work are crucial because they can assess the various critical steps of this procedure. In this work, we show through expert feedback that our developed metrics are important. Moreover, we demonstrate the discriminant validity of our developed task-specific metrics by conducting a study using a porcine small bowel model where participants performed side-by-side anastomosis using a linear stapler. We further improved the sensitivity of our metrics using the PLS regression technique and identified essential metric items that are critical in assessing the level of expertise.

The survey to assess the importance of the developed metrics showed that the majority of the less than 5 years group rated all the metric items very important or important, whereas the 5 to 15 years group had been more critical and rated some of the metric items as less important. The more than 15 years group's rating was between the other 2 groups. Statistical analysis showed that except for 2 metric items, there was a general consensus. In our study, our choice of the model for the anastomosis task was a porcine small bowel which was bought frozen with intestines precleaned and thawed in the required quantity each day. The porcine small bowel provided a real tissue-like feel compared with silicone anastomosis models and enabled us to perform a leak test. When assessing performance from the recorded videos, we omitted some of the metrics because they are not relevant for the porcine small bowel model; specifically, metrics related to bleeding, alignment of the small bowel (which was already set by the experimenter), the metric for assessing the size, and the General metrics that are part of the Global metrics. After rating 5 videos, we combined certain metric items to improve consistency and better agreement between our 2 raters. For example, metric items M11 (open enterotomy point) and M12 (grasp enterotomy point on pieces of the intestine) were combined to M11-12 (grasp and excise enterotomy point on pieces of intestines).

Analysis of results showed that our task-specific metrics were able to differentiate performance between our novice and experienced groups. Correlation analysis showed a moderate correlation to the global metric item. After PLS regression, the correlation improved

significantly to the global score. The PLS weights also showed that some metric items were weighted higher with the highest weight for metric M11-12. Three metric items, M24 (push the tab completely down to staple and cut), M27 (suture handling during oversewing of transverse staple line), and M35 (overall suture handling) had negative weights, indicating that those assessment items contributions need to be reduced to obtain maximum correlation with the global score. In our study, we noticed that 3 experienced operators and 2 novices did not oversew the staple line and 1 experienced operator oversewed only the edge. However, oversewing offers protection against anastomosis leak and reoperative surgery rate.[38] We also noticed that only 1 of the 9 subjects analyzed in the novice group performed the apical stitch compared with 6 in our experienced group. The metric M11-12 showed the largest variation, with 7 experienced operators getting perfect scores compared with the novice group, where none got a perfect score. We also observed that because it is the first step in our study, novice participants with no previous experience had difficulty figuring out what to do. The PGY1s in our group had the most difficulty, with 2 subjects starting to suture the 2 ends together and 1 trying to insert the stapler without making the incision. Metric reduction analysis using PLS also showed that this metric item was the most important in predicting the performance.

During the leak test, there was no clear difference in leak rate between the 2 groups. In side-to-side anastomosis, it has been observed that for 2 rows of staples, both side and apex of the staple line are weak points.[39,40] Because many of our participants either did not perform an apical stitch or did that poorly, the leak rate is not surprising. Another contributing factor could be attributable to the time elapsed between the anastomosis and testing. We collected all the specimens and performed the test in batches, sometimes 2–3 days after completion of the task. During that time, the specimens were stored in a refrigerator and thawed before testing, which may lead to deterioration in the specimens' biomechanical properties such as elasticity. Moreover, in frozen and thawed specimens, some of the properties of live tissues may be lost, including the ability to close suture holes. A similar study that used bursting pressure for assessment of junior and experienced group performance on stapled side-by-side anastomosis on a porcine model also observed a large variation in bursting pressure and determined it was not reliable in assessing the performance.[41]

There are several limitations to our study. Because of the use of an explanted porcine small bowel, not all the metrics developed were assessed in this study. Because the goal of this study is to assess the technical performance in basic linear stapled–based small bowel anastomosis, we did not assess antiperistaltic/isoperistaltic techniques which are important but were outside the scope of this work. The closing of mesenteric defect is important but was not assessed in this work owing to the limitations of our porcine small bowel model. When assessing the combined metric item M11-12, our intention was to check whether the bowel was grasped in full thickness (mucosa to serosa) using the Allis forceps, but owing to the placement of the digital camera to capture the entire task, it could not be used to assess full thickness bite. Instead, we used the position of the bite (how far from the edge) to assess it. Furthermore, in this study, the experimenter loaded the stapler every time after firing and didn't test whether the subjects could properly assemble the stapler, reload staples, or assess it before using, which could be an important skill for assessment.

CONCLUSIONS

Task-specific metric items were developed based on expert consensus and showed good discriminant validity when compared with a global rating scale to discriminate between experienced and novice operators. These metrics can be used for evaluating technical skills in a stapled small bowel anastomosis model. Our next step is to incorporate these metrics in our VCOST simulator and validate it.

Author Contributions

Conceptualization: De, Fleshman, Sankaranarayanan, Halic
Data curation: Sankaranarayanan, Parker, Demirel, Halic, Jacinto
Formal analysis: Sankaranarayanan, Demirel, Halic, Parker, De,
Funding acquisition: De, Sankaranarayanan, Halic, Fleshman
Investigation: Sankaranarayanan, Fleshman, Parker, Halic, Jacinto, De
Methodology: Sankaranarayanan, Fleshman, Demirel, Halic
Project administration: De, Sankaranarayanan, Fleshman
Resources: De, Fleshman
Supervision: De, Fleshman
Writing – original draft: Sankaranarayanan, Demirel, Halic, Parker, Jacinto, De, Fleshman
Writing – review & editing: Sankaranarayanan, Parker, Jacinto, Demirel, Halic, De, Fleshman

Acknowledgment The authors thank Catherine Kumwenda-Wilson for her help in running the study.

REFERENCES

1. Williams RG, Kim MJ, Dunnington GL. Practice guidelines for operative performance assessments. Ann Surg 2016;264:934–948.
2. Tsuda S, Scott D, Doyle J, Jones DB. Surgical skills training and simulation. Curr Probl Surg 2009;46:271–370.
3. Woodrum DT, Andreatta PB, Yellamanchilli RK, et al. Construct validity of the LapSim laparoscopic surgical simulator. Am J Surg 2006;191:28–32.
4. Fraser SA, Klassen DR, Feldman LS, et al. Evaluating laparoscopic skills, setting the pass/fail score for the MISTELS system. Surg Endosc Other Interv Tech 2003;17:964–967.
5. Vassiliou MC, Dunkin BJ, Marks JM, Fried GM. FLS and FES: Comprehensive models of training and assessment. Surg Clin North Am 2010;90:535–558.
6. Ritter EM, Scott DJ. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. Surg Innov 2007;14:107–112.
7. Deal SB, Stefanidis D, Telem D, et al. Evaluation of crowdsourced assessment of the critical view of safety in laparoscopic cholecystectomy. Surg Endosc Other Interv Tech 2017;31:5094–5100.
8. Jung JJ, Jüni P, Gee DW, et al. Development and evaluation of a novel instrument to measure severity of intraoperative events using video data. Ann Surg 2020;272:220–226.
9. Goldenberg MG, Grantcharov TP. Video-analysis for the assessment of practical skill. Tijdschr voor Urol 2016;6:128–136.
10. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med 2013;369:1434–1442.
11. Hogg ME, Zenati M, Novak S, et al. Grading of surgeon technical performance predicts postoperative pancreatic fistula for pancreaticoduodenectomy independent of patient-related variables. Ann Surg 2016;264:482–489.
12. ABS to Explore Video-Based Assessment in Pilot Program Launching June 2021 | American Board of Surgery. Available at: https://www.absurgery.org/default.jsp?news_vba04.21. Accessed January 25, 2022.
13. Pugh CM, Hashimoto DA, Korndorffer JR. The what? how? and who? of video based assessment. Am J Surg 2021;221:13–18.
14. Demirel D, Butler KL, Halic T, et al. A hierarchical task analysis of cricothyroidotomy procedure for a virtual airway skills trainer (VAST) simulator. Am J Surg 2016;212:475–484.
15. Sankaranarayanan G, Cetinsaya B, Halic T, et al. Development of structured objective metrics for automated assessment of double layered end-to-end small bowel anastomosis. Dis colon Rectum 2020;63:E377.
16. Abdelsattar JM, AlJamal YN, Ruparel RK, et al. Correlation of objective assessment data with general surgery resident in-training evaluation reports and operative volumes. J Surg Educ 2018;75:1430–1436.
17. Feldman LS, Hagarty SE, Ghitulescu G, et al. Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. J Am Coll Surg 2004;198:105–110.
18. De Montbrun SL, Roberts PL, Lowry AC, et al. 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 2013;258:1001–1006.
19. Parker LM, Khan A, Kumwendawilson C, et al. Validation of task-specific metrics for the assessment of hand-sewn bowel anastomoses: Developing a virtual reality-based colorectal surgical trainer. Dis Colon Rectum 2021;64:E377.
20. de Montbrun S, Roberts PL, Satterthwaite L, MacRae H. Implementing and evaluating a national certification technical skills examination: The colorectal objective structured assessment of technical skill. Ann Surg 2016;264:1–6.
21. de Montbrun SL, Roberts PL, Lowry AC, et al. A novel approach to assessing technical competence of colorectal surgery residents. Ann Surg 2013;258:1001–1006.
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995;57:289–300.
23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–163.
24. Frank IE, Friedman JH. A Statistical view of some chemometrics regression tools. Technometrics 1993;35:109–135.
25. Höskuldsson A. PLS regression methods. J Chemom 1988;2:211–228.
26. Abdi H. Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdiscip Rev Comput Stat 2010;2:97–106.
27. Ryason A, Petrusa ER, Kruger U, et al. Development of an endotracheal intubation formative assessment tool. J Educ Perioper Med 2020;22:E635–E635. Available at: https://pubmed.ncbi.nlm.nih.gov/32432150. Accessed September 7, 2022.
28. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965;52:591–611.
29. R Core Team. R: A language and environment for statistical computing. Vienna, Austria. 2020.
30. Clatterbuck B, Moore L. Bowel, Small Resection. StatPearls Publishing; 2018. Available at: https://www.ncbi.nlm.nih.gov/books/NBK507896/. Accessed March 16, 2022.
31. Korolija D. The current evidence on stapled versus hand-sewn anastomoses in the digestive tract. Minim Invasive Ther Allied Technol 2008;17:151–154.
32. Dziki AJ, Ph D, Duncan MD, et al. Advantages of hand-sewn over stapled bowel anastomosis. Dis Colon Rectum 1991;34:442–448.
33. Hanson RR, Nixon AJ, Calderwood-Mays M, et al. Comparison of staple and suture techniques for end-to-end anastomosis of the small colon in horses. Am J Vet Res 1988;49:1621–1628.
34. Slieker JC, Daams F, Mulder IM, et al. Systematic review of the technique of colorectal anastomosis. JAMA Surg 2021;148:190–201.
35. Brown SL, Woo EK. Surgical stapler-associated fatalities and adverse events reported to the food and drug administration. J Am Coll Surg 2004;199:374–381.
36. Lustosa SADS, Matos D, Atallah ÁN, Castro AA. Stapled versus handsewn methods for colorectal anastomosis surgery:

A systematic review of randomized controlled trials. Sao Paulo Med J 2002;120:132–136.
37. Betzold R, Laryea JA. Staple line/anastomotic reinforcement and other adjuncts: Do they make a difference? Clin Colon Rectal Surg 2014;27:156–161.
38. Baqar AR, Wilkins S, Staples M, et al. The post-operative impact of oversewing stapled anastomoses in colorectal cancer surgery: A retrospective Australian cohort study. Int J Surg Open 2020;24:91–95.
39. Kimura M, Takahashi H, Tanaka T, et al. Weak points of a stapled side-to-side anastomosis. Hepatogastroenterology 2015;62:924–926.
40. Kimura M, Kuwabara Y, Taniwaki S, et al. Improving the side-to-side stapled anastomosis: comparison of staplers for robust crotch formation. Surg Obes Relat Dis 2018;14:16–21.
41. Vick LR, Vick KD, Borman KR, Salameh JR. Face, content, and construct validities of inanimate intestinal anastomoses simulation. J Surg Educ 2007;64:365–368.