Surgical Endoscopy (2023) 37:1282–1292
https://doi.org/10.1007/s00464-022-09584-1

ORIGINAL ARTICLE

Objective metrics for hand-sewn bowel anastomoses can differentiate novice from expert surgeons

Ganesh Sankaranarayanan[1] · Lisa M. Parker[2] · Aimal Khan[3] · James Dials[4] · Doga Demirel[4] · Tansel Halic[5] · Alyson Crawford[6] · Uwe Kruger[7] · Suvranu De[7] · James W. Fleshman[2]

Received: 13 February 2022 / Accepted: 25 August 2022 / Published online: 30 September 2022

Abstract

Background Assessing performance automatically in a virtual reality trainer or from recorded videos is advantageous but needs validated objective metrics. The purpose of this study is to obtain expert consensus and validate task-specific metrics developed for assessing performance in double-layered end-to-end anastomosis.

Materials and methods Subjects were recruited into expert (PGY 4–5, colorectal surgery residents, and attendings) and novice (PGY 1–3) groups. Weighted average scores of experts for each metric item, completion time, and the total scores computed using global and task-specific metrics were computed for assessment.

Results A total of 43 expert surgeons rated our task-specific metric items with weighted averages ranging from 3.33 to 4.5 on a 5-point Likert scale. A total of 20 subjects (10 novices and 10 experts) participated in validation study. The novice group completed the task significantly more slowly than the experienced group (37.67 ± 7.09 vs 25.47 ± 7.82 min, $p = 0.001$). In addition, both the global rating scale (23.47 ± 4.28 vs 28.3 ± 3.85, $p = 0.016$) and the task-specific metrics showed a significant difference in performance between the two groups (38.77 ± 2.83 vs 42.58 ± 4.56 $p = 0.027$) following partial least-squares (PLS) regression. Furthermore, PLS regression showed that only two metric items (Stay suture tension and Tool handling) could reliably differentiate the performance between the groups (20.41 ± 2.42 vs 24.28 ± 4.09 vs, $p = 0.037$).

Conclusions Our study shows that our task-specific metrics have significant discriminant validity and can be used to evaluate the technical skills for this procedure.

Technical competency of surgeons directly correlates with their surgical outcomes and rate of complications [1]. Currently, there is no direct assessment of surgical trainees' technical skills before entering independent surgical practice. Trainees in various surgical specialties (including Colon and Rectal Surgery) are tested using only oral and written board exams to ensure that they have the appropriate knowledge to function as competent surgeons, without any objective assessment of their technical skills. The only assessment of technical skills is performed by individual training programs using subjective faculty evaluations. Studies comparing these evaluations to more objective assessments (like task-specific metrics) have found they do not evaluate a residents' technical performance as well as the objective evaluations [2, 3].

Different surgical specialties have attempted developing structured assessment tools to assess the technical skills of their trainees. These tools have involved a combination of task-specific metrics, global rating scales, time for the procedure, and a pass/fail grade by the level of training. They have been able to assess a trainee's surgical skills reliably and with good validity [4]. 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

Presented as podium at the Virtual ASCRS 2021 conference April 24–28, 20201.

✉ Ganesh Sankaranarayanan
Ganesh.Sankaranarayanan@UTSouthwestern.edu

1 Department of Surgery, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390, USA

2 Baylor University Medical Center, Dallas, TX, USA

3 Vanderbilt University Medical Center, Nashville, TN, USA

4 Florida Polytechnic University, Lakeland, FL, USA

5 University of Central Arkansas, Conway, AK, USA

6 Western Surgical Group, Reno, NV, USA

7 Rensselaer Polytechnic Institute, Troy, NY, USA

address the question of assessing technical competency of 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 were able to reliably differentiate colorectal surgery residents from general surgery residents [5]. Despite their merits, skill assessment tools require a considerable amount of time, labor, and materials to set up for each participant. Additionally, a proctor is needed to administer and score the performance which adds additional barrier to participation.

As a potential alternative, virtual reality-based surgical simulators offer several advantages including objective scoring, quick turnaround times, and no need to replenish materials. Virtual reality-based simulators have been widely applied for training in various laparoscopic [6–10], robotic [11–14], and endoscopic procedures [15, 16], However, the application of this technology to open surgery remains very limited [17]. We are developing a virtual reality-based colorectal surgery trainer (VCOST) for assessment and training in select open colorectal surgical tasks. As an integral part of many colorectal procedures, our first focus is to develop a virtual reality-based simulator for hand-sewn bowel anastomosis [18]. Using expert consensus, task-specific objective metrics were developed for automated assessment of double-layered hand-sewn end-to-end anastomosis. Here, we test the ability of these task-specific metrics to differentiate between novice and expert surgeons.

Materials and methods

Development of task-specific metrics for bowel anastomosis

In this IRB-approved study, a detailed Hierarchical Task Analysis (HTA) was created based on observing expert surgeons performing a double-layered end-to-end hand-sewn anastomosis on porcine bowel. Two expert colorectal surgeons (AC, JF) recorded videos for use which were supplemented with an extensive literature review and publicly available workshop videos. This resulted in the documentation of the major tasks and subtasks of the procedure in the HTA. A performance metric was associated with each significant task related to surgical action. We used a 5-point scale for the performance metrics, with five being wholly correct and 0 being completely incorrect. Consensus on the importance of the developed metrics and the preference of suturing techniques and materials was obtained from expert colorectal surgeons using an online survey on a 5-point Likert scale.

Validation of the metrics

Once these metrics were developed, a study was conducted at the Baylor University Medical Center (BUMC) in Dallas to assess the discriminant validity of the metrics in differentiating between expert and noviceparticipants. The participants were asked to perform a double-layered end-to-end anastomosis on a small bowel silicone model (Simsei, Applied Medical Inc.). Since our goal is to validate the developed task-specific metrics, apart from an explanation of the setup and the model, no additional information such as written description or an expert demonstration video was provided to the participants about the task. The 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 three years of their training were assigned to the novice group. Some of them may have been exposed to colorectal procedures as part of their trauma and critical care rotation, but we do not specifically control for it in this study. The participants were deidentified and videotaped performing a bowel anastomosis. A digital camera and an Intel Realsense 3D-depth camera were used to record the procedures. Additionally, the positions of the tools and the surgeon's hands were tracked using magnetic tracking sensors (3D Guidance Ascension TrakStar). The tool and hand motion data were not used in this analysis but will be used to analyze the workspace for future development of a custom haptic device for our simulator. The depth data was collected for future use and was not used for analysis in this study. The experimental setup is shown in Fig. 1. Two qualified independent raters assessed the videos. The rating on the recorded video was done using the developed task-specific metrics and the global rating scale of performance previously developed and validated for the COSATS [19, 20]. To assist the raters in assessing the quality of the anastomosis, photos were taken of both sides of the exterior and the interior of the models in which anastomosis were performed.

Statistical analysis

For the analysis of the data from the survey, descriptive statistics were used, and a weighted average was calculated for all the metric items.

For the analysis of data from the validation study, Intraclass Correlation Coefficient (ICC) was computed for absolute agreement to assess the inter-rater reliability of the two raters. The ICC was calculated after the completion of the first five videos by the raters. Discrepancies were resolved, and the process was repeated for the second

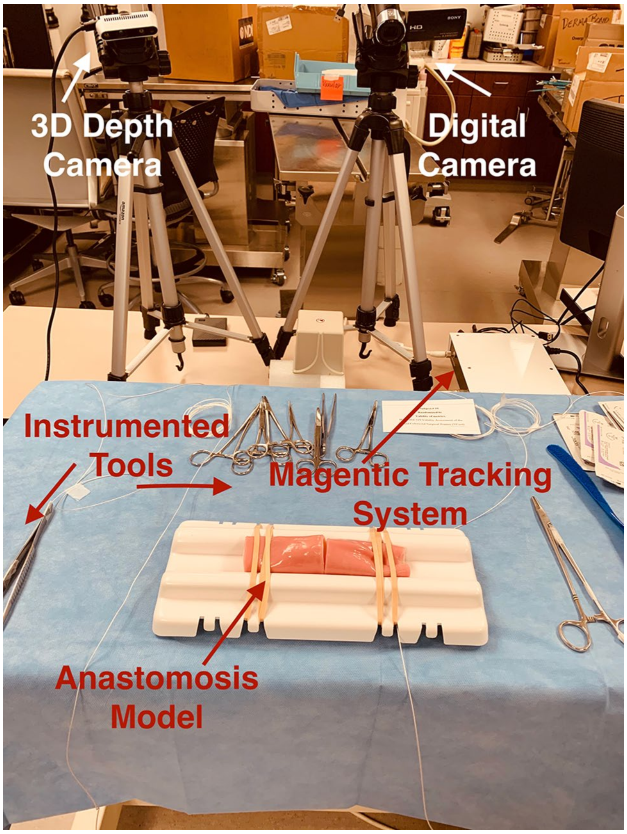

Fig. 1 Experimental setup for the validation of the developed metrics for the hand-sewn anastomosis

set of 5 videos. On completion of the ratings, we averaged the ratings from the two raters to calculate the global and task-specific scores.

For the global metrics, the total score was computed by adding all the metric items. For the task-specific metric score, two methods were used. In the first method, a total task-specific metric score was computed by adding all the individual metrics. In the second method, a weighted sum was computed with the weights for individual task-specific metric items calculated using Partial Least Squares (PLS) regression [21] for predicting the total global metrics score. PLS is a well-known statistical method that consists of a latent variable approach to model the covariance structure between the predictor (task-specific metrics) and the response (total global score) variables. It is particularly useful if only a relatively small dataset is available [22]. The PLS regression technique has been successfully used in developing a formative assessment tool for Endotracheal Intubation [23]. In this work, an independent cross-validation method was used to evaluate the PLS model. Additionally, a Support Vector Machine (SVM), a supervised machine learning method, was used to validate the results from PLS regression independently. Finally, the completion time of each participant was also recorded from the videos for analysis.

For completion time, and total Global and Checklist scores, the normality of data was checked using a Shapiro–Wilk test [24]. If the data were normal, we used the t test to compute the difference between the groups; otherwise, we used the non-parametric Mann–Whitney test. All statistical analyses, except the SVM, were performed using the R statistical package (version 4.0.2) [25]. The SVM was computed using custom code implemented in Matlab (Mathworks Inc.).

Results

Hierarchical task analysis results

The HTA was conducted to identify the key steps of the procedure and represent them in a hierarchical order. The task tree (Fig. 2) was created to determine the optimal execution of the steps and substeps, including their linear progression. HTA was then used to develop the performance metrics. Six distinct main steps were identified in the HTA; 1) Placement of stay sutures, Completion of suturing of the 2) Posterior Outer, 3) Posterior Inner, 4) Anterior Inner, 5) Anterior Outer Layer and 6) Evaluation steps. The details of each of these steps is shown in Table 1.

Expert consensus survey results

Most of the surgeons (55.8%) indicated double-layer closure as their preferred method while 25.5% preferred single layer, and 18.6% didn't indicate any preference. When asked about the importance of time in the execution of the task, 51.5% indicated it was important, 21.2% didn't think it was important with 3% indicated it was very important and 24.2% remained neutral. When asked how much time is adequate to perform this task, 60.6% indicated that 10–30 min is an adequate time to perform the double-layer end-to-end hand-sewn anastomosis. Vicryl was the most preferred suture material to close the inner wall, followed by PDS and Chromic (58.84%, 32.26%, 6.45%). The outer wall closure was most commonly done using Vicryl (48.48%), followed by Silk (45.45%) and PDS (6.06%).

A majority of the surgeons (66.67%) chose Connell stitch as their preferred suturing method for the inner layer, while 81.82% used Lembert sutures for the outer layer. Additionally, 71.43% of the surgeons responded that they start in the middle and run in either direction when closing the inner layer of the posterior wall. For distances to the first bite from the cut end of the bowel, 55.88% preferred 5 mm from the cut edge while 35.29% preferred < 5 mm from the cut edge. The metrics for assessment and their weighted average

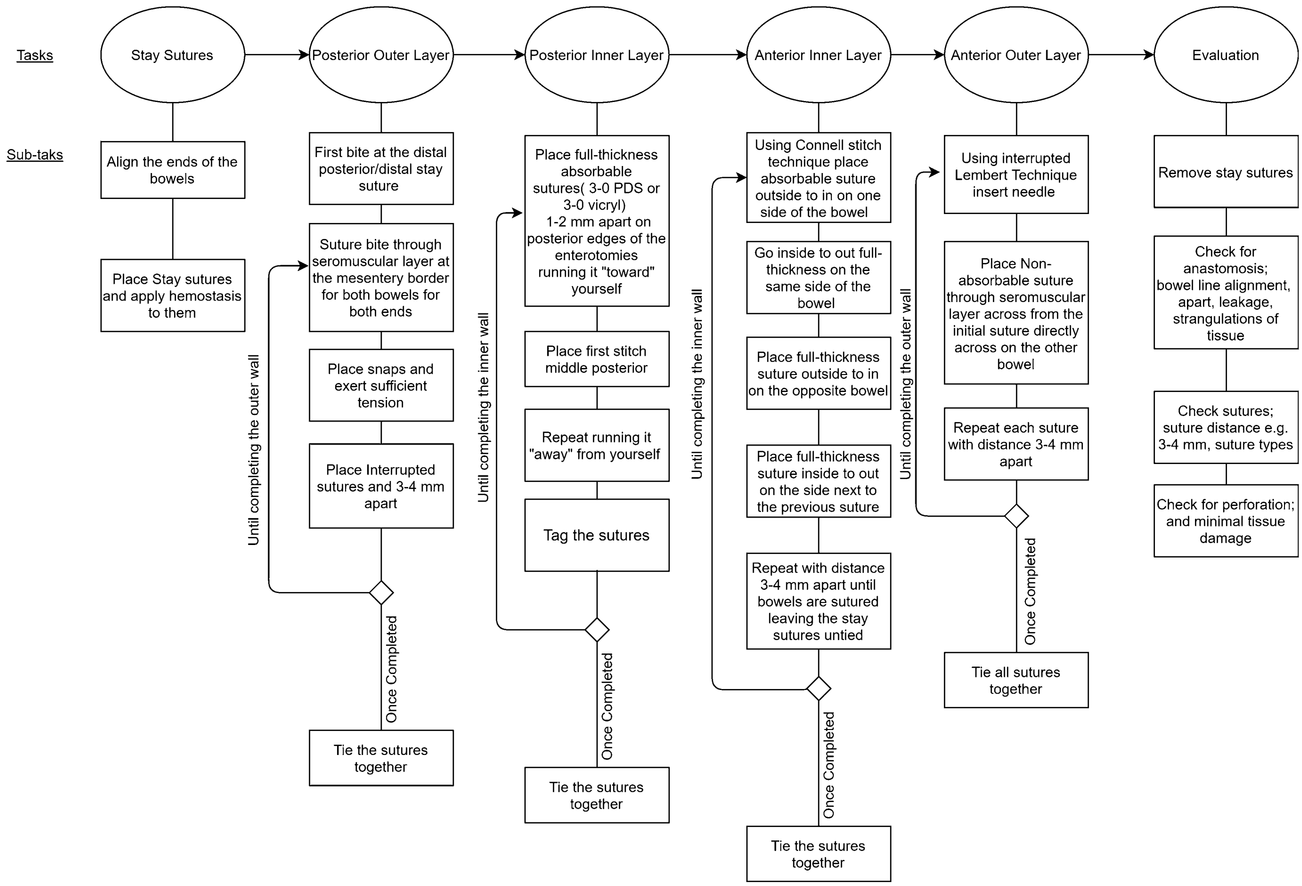

Fig. 2 Task tree showing the task execution steps of the procedure in order

scores are shown in Table 2 and ranged from 3.33 (above neutral) to 4.5 (more than important).

Validation of metrics results

To validate the developed task-specific metrics, a total of ($n = 20$) subjects distributed equally between the expert and novice groups participated in this study. Figure 3 shows a participant performing the task during the study. The anterior, posterior, and the inner view of the completed anastomosis model for one subject is shown in Fig. 4. The completion time for all the subjects is shown in Fig. 5. The novice group took more time (37.67 ± 7.09 min) compared to the expert group (25.47 ± 7.82 min). T test results show that the difference in completion time was significant ($p = 0.001$).

Reliability analysis

Intra class correlation (ICC) analysis of ratings by the two raters showed that the global rating scale had moderate reliability (ICC = 0.64,95%CI 0.5 – 0.7, $p < 0.0001$) and the task-specific metrics had good reliability (ICC = 0.801, 95%CI 0.72 – 0.85, $p < 0.0001$).

Global score metrics

The total global metric score computed from the average of the two raters is shown in Fig. 6. The expert group scored significantly higher in total global scores than the novice group (28.3 ± 3.85 vs 23.47 ± 4.28, $p = 0.016$).

Task-specific metrics

For the task-specific metric, seven metric items from a total of 10 items (see Table 2) were used. The metrics for anastomosis quality check, closure of the mesenteric defect, and management of bleeding were not used since we used a silicone-based double-layered small bowel model. The total task-specific metric scores, computed from adding all seven metric items, are shown in Fig. 7. The expert group had higher overall task specifc metric scores than the novice group (26.05 ± 4.67 vs 23.75 ± 4.67, $p = 0.7$), but it was not statistically significant. The Spearman correlation coefficient

Table 1 The main operative steps in performing double-layered end-to-end hand-sewn anastomosis

Step	Description
(1) Placement of stay sutures	The procedure starts with aligning both bowels end-to-end and placing absorbable 3–0 silk stay sutures. An initial stay suture is placed distal and proximal 5 mm from the cut edge of the intestine at the posterior outer layer. Hemostats are applied to each stay suture and adjusted to ensure adequate tension on the bowel to line up the cut edges
(2) Posterior outer layer suturing	After securing the stay sutures with hemostats, the posterior outer layer of anastomosis is created using 3–0 silk in an interrupted lambert fashion incorporating only the serosa and muscularis with each bite (partial thickness). The surgeon needs to place a hemostat on the sutures and leave them untied with enough tension until all the sutures are placed and then tie them in the end. The interrupted sutures are placed 3-4 mm distance apart from each other [26, 27]. Once the outer wall is completed, the sutures are tied together (e.g., batching technique)
(3) Posterior inner layer suturing	For the posterior inner layer, a running suture is placed using a 3–0 PDS or Vicryl absorbable sutures taking 1–2 mm bites on posterior edges of the enterotomies [28, 29]. For this portion two such sutures are used starting in the middle and sutured away from the middle of the anastamosis. These two sutures are also tied together
(4) Anterior inner layer suturing	Using the Connell stitch technique, the surgeon brings the 3–0 PDS or 3–0 Vicryl absorbable sutures of the posterior inner layer on both sides to to the anterior wall. Leaving the stay sutures untied, the sutures from both end are brought towards the middle of the anterior wall with with 3-4 mm bites until both bowels are sutured and the inner wall is completely sutured. The two sutures are then tied together
(5) Anterior outer layer suturing	For the anterior outer layer, using the interrupted Lembert technique, non-absorbable 3–0 silk sutures are placed through the seromuscular layer across from the initial suture on the other bowel. Leaving the Lembert sutures untied, similar to the inner layer, the same pattern is repeated, leaving 3-4 mm distance apart from each other until both bowels are sutured over the outer wall. Batching, tying all the sutures at the end as opposed to the individual ties for each suture, is optimally used here as well
(6) Evaluation	The stay sutures are removed, the bowel, bowel line alignment, perforation, leakage, strangulations of the tissue are examined. The suture types used and the distance are confirmed as a part of the assessment

between the total global score and the total task-specific metric scores (Fig. 8), showed poor correlation with $R = 0.483$, $p = 0.03$ and $R^2 = 0.25$.

The weights for each of the task-specific metric item computed using the PLS regression is shown in Table 3. The weights varied for different metric items. Notably, the weight for stay suture tension was negative, which was because some of the participants, including experts, didn't place a stay suture or provided adequate tension since we had the models secured with rubber bands on a plastic suturing platform. A new total task metric score was then computed using the weights in Table 3. Figure 9 shows the correlation between the total global score and the total task metric score computed after the PLS regression using all the metric items. The correlation improved significantly with $R = 0.901$, $p < 0.0001$ and $R^2 = 0.764$. Once the new task-specific metric scores were computed using PLS regression, the performance between the groups (Fig. 10) showed that the experts performed significantly better than the novices (42.58 ± 4.56 vs 38.77 ± 2.83, $p = 0.027$).

Task-specific metric reduction

We did a systematic study to discover the minimal set of procedure-specific metric items that could discriminate between experts and novices. To do this, we performed the PLS regression by leaving out one metric item at a time to compute a new total score and repeated it for all the seven metric items. Once the new scores were obtained using the computed PLS weights, we then performed the comparison between the two groups. Table 4 shows the results from this analysis. For brevity, we only report the R^2 value from the correlation between the global score and the reduced task-specific score and the t test results.

Table 4 shows that two metric items—tool handling and the stay suture tension—are critical in predicting the performance between the two groups since removing them made our computed task-specific metric unable to differentiate between the groups.

Next, we computed a new task-specific metric score by running the PLS regression model using only the reduced set of two metric items (tool handling and stay suture tension). The weights for the two task-specific metric items computed using the PLS regression are shown in Table 5. The tool handling metric item weight is 18 times more than the stay suture tension metric, indicating that it has significantly more influence on the assessment of performance. As described before, because we had secured our anastomosis model using rubber bands, many participants did not place proper tension on the stay suture and hence that metric item, though important, contributed only slightly to the assessment of the performance.

Figure 11 shows the correlation between the total global score and the total task metric score computed after the

Table 2 Task-specific metric items and the weighted average scores from the survey

Number	Metrics	Agreement (weighted average of 5-point likert scale)
1	Task execution order a. Suturing of outer bowel wall followed by suturing of the inner bowel wall (5 points) b. Suturing of inner bowel wall followed by suturing of the outer wall (0 point)	3.79
2	Suture handling a. Equidistant placement of sutures (5 points) b. Poor placement of sutures (3 points) c. Inadequate suture handling (0 point)	4.53
3	Tool selection a. Proper selection of surgical tools (5 points) b. Proper selection of tools some of the time (3 points) c. Improper selection of tools (0 points)	4.13
4	Tool handling a. Smoothness and Gentleness in tool handling (5 points) b. Discrete motions in tool handling (3 point) c. Aggressive tool handling (0 point)	4.34
5	Anastomosis quality check Checking for integrity and leakage (air or betadine) a. No leak (5 points) b. Leaks (0 point)	3.52
6	Intestinal tissue damage a. No damage (5 points) b. Damage due to any reason, instrument use etc. (0 point)	4.48
7	Managing bleeding during the procedure a. Immediately (5 points) b. Delayed (0 point)	4.18
8	Closure of the mesenteric defect a. Completely closed (5 points) b. Not closed (0 point)	2.48
9	Placement of stay sutures a. Distal and proximal stay sutures (5 points) b. No stay sutures placed (0 point)	3.42
10	Stay sutures tension a. Placing snaps and having enough tension (5 points) b. Not placing snaps or not having enough tension (0 point)	3.50

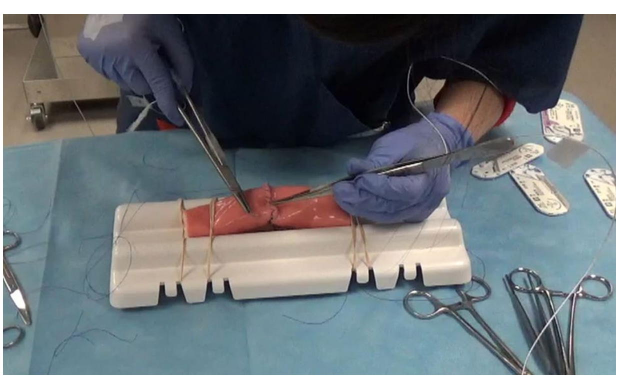

Fig. 3 Experiment setup showing the anastomosis model and a subject performing the task

PLS regression on the reduced set of metrics. There was a fair correlation with R = 0.856, $p < 0.0001$ and $R^2 = 0.671$. Figure 12 shows the plot of the scores with a reduced set of metrics between the two groups. Mann–Whitney U test shows that the expert group performed better than the novice group (24.28 ± 4.09 vs 20.41 ± 2.42, $p = 0.037$).

We also independently verified the importance of the two metric items using an SVM analysis where the combination of our metric item provided the highest minimum classification error (MCE = 0.8).

Discussion

Multiple techniques may be used to perform a bowel anastomosis task. With the advent of staplers, the overall experience of hand-sewn bowel anastomoses, in particular, is

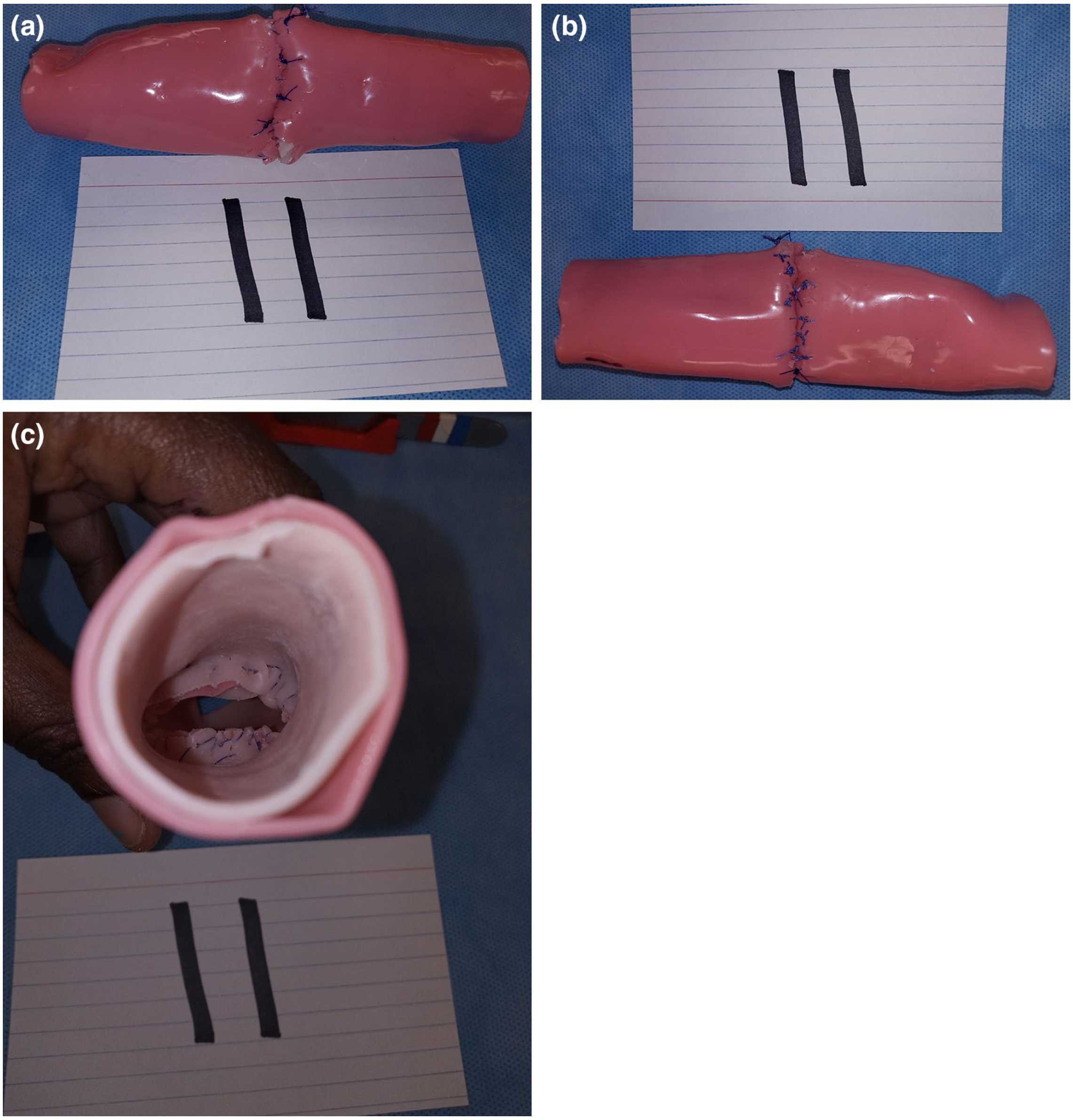

Fig. 4 **a** Anterior, **b** Posterior, **c** Inner view of the completed silicone anastomosis model

declining in general surgery residencies [31]. As a trained colorectal surgeon, there are circumstances, such as a low coloanal anastomosis or a very proximal small bowel anastomosis, where a hand-sewn anastomosis would be the optimal, or only, technique feasible. Furthermore, along with patient and situational factors, a poorly constructed bowel anastomosis can result in an anastomotic leak, leading to significant morbidity and mortality [32]. An objective structured assessment tool could be utilized to ensure that this level of expertise is reached after adequate training.

Training in hand-sewn anastomosis task is usually provided to all general surgery residents by their third year or before entering their colorectal surgery rotation. The training is largely based on benchtop synthetic, silicone, porcine, or cadaveric bowel models with and without leak tests. It has been shown that training in benchtop models has shown improvement in skills as assessed by OSATS, decrease in operating time, and better anastomosis quality as assessed by leak test [33, 34].

Our results indicate that our task-specific metrics can objectively differentiate between novice and experts performing hand-sewn bowel anastomoses. Using the PLS regression analysis, we have also shown how to apply formative assessment metrics in a systematic way to assess technical skills. Moreover, we have shown that with the reduced

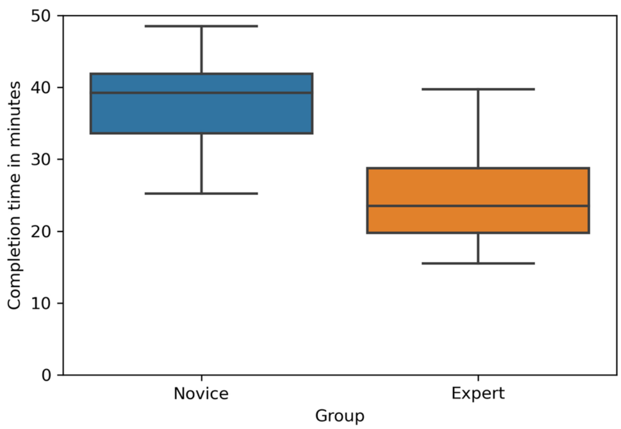

Fig. 5 Completion time in minutes for both the groups

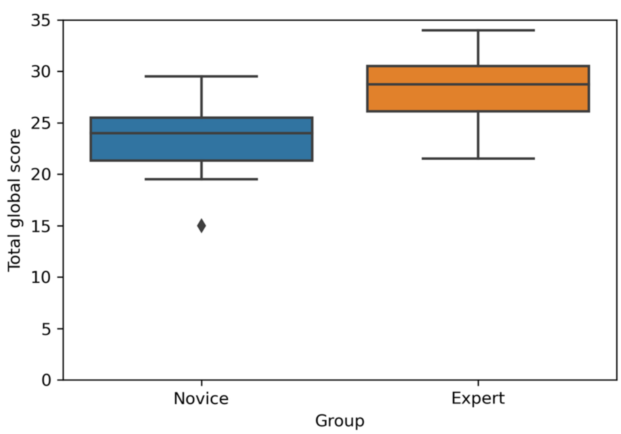

Fig. 6 Total global score for both the groups

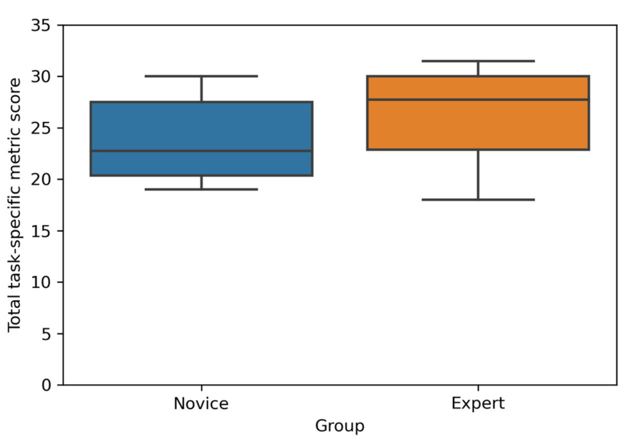

Fig. 7 Total task-specific score for both the groups

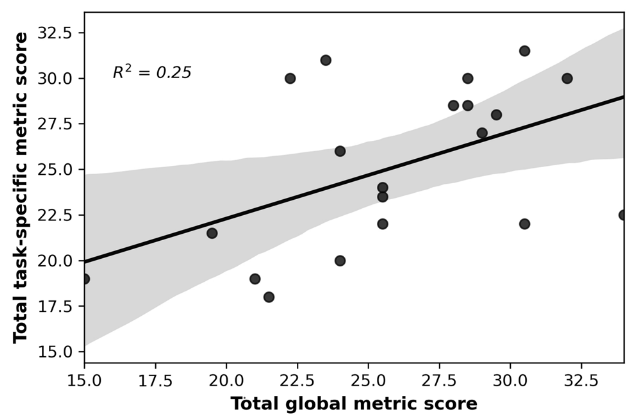

Fig. 8 Correlation between the total global score and total task-specific metric score and a linear regression model fit with 95% confidence band

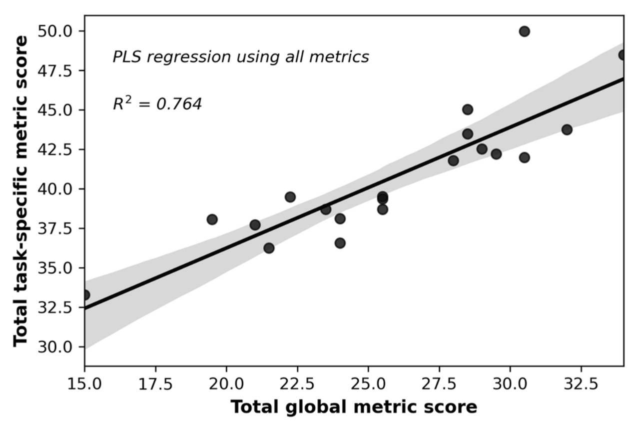

Fig. 9 Correlation between the total global score and total task-specific metric score computed using the PLS weight and a linear regression model fit with 95% confidence interval

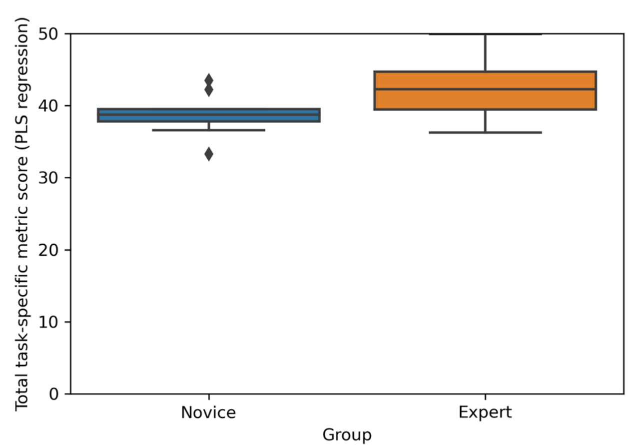

Fig. 10 Total task-specific score for both the groups computed using the PLS regression

Table 3 Weights computed for each of the metric items in task-specific metrics

Task metric items	Task execution order	Suture handling	Tool selection	Tool handling	Placement of stay sutures	Stay suture tension	Intestinal tissue damage
Weights Computed using the PLS regression	0.13025	0.98778	2.83571	5.24869	0.95602	–0.73074	0.9892

Table 4 Results from the task-specific metrics computed using the PLS regression with one metric item removed at a time

Metric item removed	Task execution order	Suture handling	Tool selection	Tool handling	Placement of stay suture	Stay suture tension	Intestinal tissue damage
R^2 (global score vs reduced task-specific score)	0.747	0.791	0.701	0.447	0.755	0.76	0.757
Expert vs novice t test (p value)	0.032	0.033	0.017	0.208	0.02	0.216	0.03

Table 5 Weights computed for the reduced set of metric items in task-specific metrics

Task Metric Items	Tool Handling	Stay Suture Tension
Weights computed using the PLS regression	6.078048	0.323394

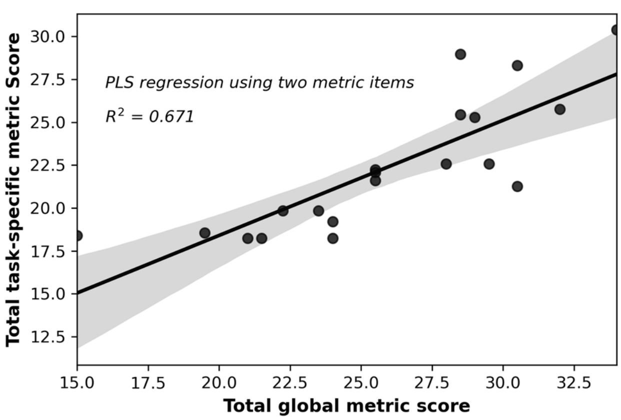

Fig. 11 Correlation between the total global score and the total task-specific metric score computed using a reduced set of two metric items

set of only two metric items, we could still differentiate the performance between the groups.

Although the use of the silicone bowel models simplified the logistics of setting up this project, there are disadvantages to using them. As the silicone models were secured, some participants, including experts, did not place stay sutures as they might in an anastomosis on real bowel. Since all metric items were equally weighted before applying the PLS regression, this could explain the discrepancy between the predictive value of the stay suture tension and the actual contribution to the total score in differentiating between the groups. Moreover, this could also explain the negative weight associated with the stay suture tension metric item. An expert who places a stay suture would put it with appropriate tension, but since some of them decided that it is unnecessary given the setup, the stay suture may not be placed at all by experts. Due to the nature of our model, we could not perform an anastomosis leak test to check the integrity of the final model as the silicone starts breaking when the knots are tightened and hence the importance of this metric in differentiating the skill level between the two groups could not be assessed. Also, we used only two raters to assess the performance from the recorded video and got only moderate agreement for the global rating scale. Having

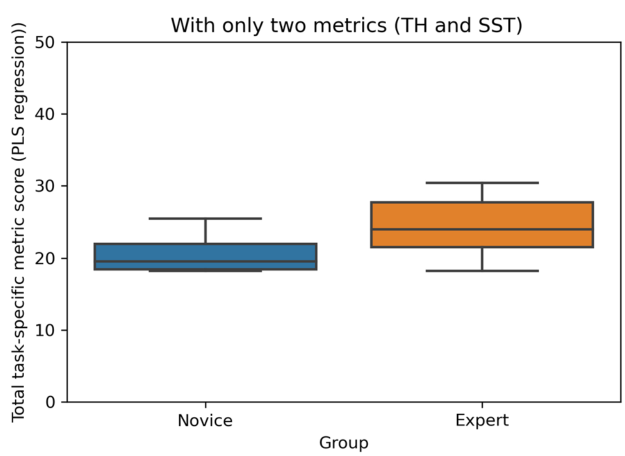

Fig. 12 Total task-specific metric score for both the groups computed using the PLS regression on the reduced set of two metric items

more than two raters could have improved agreement and overall power of the results.

Although the previously developed global metrics also differentiated between novice and experts, our task-specific metrics can be more readily applied to creating a virtual trainer for hand-sewn anastomoses. Our next step is to incorporate these metrics into VCOST and perform validation. We plan to use the methods describe in this work to develop and validate metrics for the other open surgical tasks of the COSATS and incorporate them into VCOST.

Acknowledgements The authors wish to acknowledge funding from NIH through the Grant NIH/NIBIB R01 EB025241 for this project. The authors also thank Catherine Kumwenda-Wilson for her help in running the study.

Declarations

Disclosures Drs. Ganesh Sankaranarayanan, Lisa Parker, Aimal Khan, Doga Dimeriel, Tansel Halic, Alyson Crawford, Uwe Kruger, Suvranu De, James Fleshman, and Mr. Dials have no conflicts of interest or financial ties to disclose.

References

1. Birkmeyer JD, Finks JF, O'Reilly A et al (2013) Surgical skill and complication rates after bariatric surgery. N Engl J Med 369(15):1434–1442. https://doi.org/10.1056/nejmsa1300625
2. Abdelsattar JM, AlJamal YN, Ruparel RK, Rowse PG, Heller SF, Farley DR (2018) Correlation of objective assessment data with general surgery resident in-training evaluation reports and operative volumes. j surg educ 75(6):1430–1436. https://doi.org/10.1016/j.jsurg.2018.04.016
3. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM (2004) Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. J Am Coll Surg 198(1):105–110. https://doi.org/10.1016/j.jamcollsurg.2003.08.020
4. Goff BA, Lentz GM, Lee D, Fenner D, Morris J, Mandel LS (2001) Development of a bench station objective structured assessment of technical skills. Obstet Gynecol 98(3):412–416. https://doi.org/10.1016/S0029-7844(01)01473-9
5. De Montbrun SL, Roberts PL, Lowry AC et al (2013) A novel approach to assessing technical competence of colorectal surgery residents: the development and evaluation of the colorectal objective structured assessment of technical skill (COSATS). Ann Surg 258(6):1001–1006. https://doi.org/10.1097/SLA.0b013e31829b32b8
6. Satava RM (1993) Virtual reality surgical simulator. Surg Endosc 7(3):203–205. https://doi.org/10.1007/BF00594110
7. van Dongen K, Tournoij E, van der Zee D, Schijven M, Broeders I (2007) Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 21(8):1413–1417. https://doi.org/10.1007/s00464-006-9188-2
8. Arikatla VS, Ahn W, Sankaranarayanan G, De S (2014) Towards virtual FLS: development of a peg transfer simulator. Int J Med Robot Comput Assist Surg. https://doi.org/10.1002/rcs.1534
9. Nemani A, Ahn W, Cooper C, Schwaitzberg S, De S (2017) Convergent validation and transfer of learning studies of a virtual reality-based pattern cutting simulator. Surg Endosc. https://doi.org/10.1007/s00464-017-5802-8
10. Wynn G, Lykoudis P, Berlingieri P (2018) Development and implementation of a virtual reality laparoscopic colorectal training curriculum. Am J Surg 216(3):610–617. https://doi.org/10.1016/j.amjsurg.2017.11.034
11. Lerner MA, Ayalew M, Peine WJ, Sundaram CP (2010) Does training on a virtual reality robotic simulator improve performance on the da vinci ® surgical system? J Endourol 24(3):467–472. https://doi.org/10.1089/end.2009.0190
12. Sethi AS, Peine WJ, Mohammadi Y, Sundaram CP (2009) Validation of a novel virtual reality robotic simulator. J Endourol 23(3):503–508. https://doi.org/10.1089/end.2008.0250
13. Perrenot C, Perez M, Tran N et al (2012) The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills. Surg Endosc Other Interv Tech 26(9):2587–2593. https://doi.org/10.1007/s00464-012-2237-0
14. Finnegan KT, Meraney AM, Staff I, Shichman SJ (2012) da vinci skills simulator construct validation study: correlation of prior robotic experience with overall score and time score simulator performance. Urology 80(2):330–335. https://doi.org/10.1016/j.urology.2012.02.059
15. Dorozhkin D, Nemani A, Roberts K et al (2016) Face and content validation of a virtual translumenal endoscopic surgery trainer (VTEST™). Surg Endosc Other Interv Tech 30(12):5529–5536. https://doi.org/10.1007/s00464-016-4917-7
16. Ritter E, McClusky D, Lederman A, Gallagher A, Smith C (2003) Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. J Gastrointest Surg 7(7):871–878. https://doi.org/10.1007/s11605-003-0032-x
17. Sankaranarayanan G, Li B, Miller A et al (2016) Face validation of the virtual electrosurgery skill trainer (VEST©). Surg Endosc Other Interv Tech. https://doi.org/10.1007/s00464-015-4267-x
18. Naumann DN, Bhangu A, Kelly M (2014) Clinical review stapled versus handsewn intestinal anastomosis in emergency laparotomy : a systemic review and meta-analysis. Surgery 157(4):609–618. https://doi.org/10.1016/j.surg.2014.09.030
19. de Montbrun S, Roberts PL, Satterthwaite L, MacRae H (2016) Implementing and evaluating a national certification technical skills examination: the colorectal objective structured assessment of technical skill. Ann Surg 264(1):1–6. https://doi.org/10.1097/SLA.0000000000001620
20. de Montbrun SL, Roberts PL, Lowry AC et al (2013) A novel approach to assessing technical competence of colorectal surgery residents. Ann Surg 258(6):1001–1006. https://doi.org/10.1097/SLA.0b013e31829b32b8
21. Frank IE, Friedman JH (1993) A statistical view of some chemometrics regression tools. Technometrics 35(2):109–135. https://doi.org/10.2307/1269656
22. Höskuldsson A (1988) PLS regression methods. J Chemom 2(3):211–228. https://doi.org/10.1002/cem.1180020306
23. Ryason A, Petrusa ER, Kruger U et al (2020) Development of an endotracheal intubation formative assessment tool. J Educ Perioper Med 22(1):E635
24. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3–4):591–611. https://doi.org/10.1093/biomet/52.3-4.591
25. R Core Team (2020) R: A language and environment for statistical computing. R A Lang Environ Stat Comput R Found Stat Comput Vienna, Austria
26. Ahuja V (2011) 14 - Open Surgical Techniques in Colorectal Cancer. In: Gearhart SL (ed) Early diagnosis and treatment of

cancer series: colorectal cancer 1st ed. Ahuja NBT-ED and T of CSCC W.B. Saunders, Saint Louis, pp 145–165
27. Burch JM, Franciose RJ, Moore EE, Biffl WL, Offner PJ (2000) Single-layer continuous versus two-layer interrupted intestinal anastomosis: a prospective randomized trial. Ann Surg 231(6):832–837. https://doi.org/10.1097/00000658-200006000-00007
28. Hunt SR, Silviera ML (2016) Anastomotic Construction. The ASCRS textbook of colon and rectal surgery. Springer International Publishing, Geneva, pp 141–160
29. Nezhat C, Nezhat F, Ambroze W, Pennington E (1993) Laparoscopic repair of small bowel and colon - a report of 26 cases. Surg Endosc 7(2):88–89. https://doi.org/10.1007/BF00704384
30. Slieker JC, Daams F, Mulder IM, Jeekel J, Lange JF (2013) Systematic review of the technique of colorectal anastomosis. JAMA Surg 148(2):190–201. https://doi.org/10.1001/2013.jamasurg.33
31. Nemeth ZH, Lazar EL, Paglinco SR et al (2016) Experience of general surgery residents in the creation of small bowel and colon anastomoses. J Surg Educ 73(5):844–850. https://doi.org/10.1016/j.jsurg.2016.04.011
32. Goulder F (2012) Bowel anastomoses: the theory, the practice and the evidence base. World J Gastrointest Surg 4(9):208–213. https://doi.org/10.4240/wjgs.v4.i9.208
33. Egle JP, Venkata S, Malladi S, Gopinath N, Mittal VK (2014) Simulation training improves resident performance in hand-sewn vascular and bowel anastomoses. J Surg Educ 72(2):291–296. https://doi.org/10.1016/j.jsurg.2014.09.005
34. Jensen AR, Wright AS, Mcintyre LK et al (2021) Laboratory-based instruction for skin closure and bowel anastomosis for surgical residents. Arch surg 143(9):852–859
35. De Montbrun S, Roberts PL, Satterthwaite L, Macrae H (2016) Implementing and evaluating a national certification technical skills examination. Ann Surg 264(1):1–6. https://doi.org/10.1097/SLA.0000000000001620