Surgical Endoscopy
https://doi.org/10.1007/s00464-021-08893-1

Hierarchical task analysis of endoscopic sleeve gastroplasty

James Dials[1] · Doga Demirel[1] · Tansel Halic[2] · Suvranu De[3] · Adam Ryason[3] · Shanker Kundumadam[4] · Mohammad Al-Haddad[4] · Mark A. Gromski[4]

Received: 29 March 2021 / Accepted: 16 November 2021

Abstract

Background Endoscopic sleeve gastroplasty (ESG) is a minimally invasive endoscopic weight loss procedure used to treat obesity. The long-term goal of this project is to develop a Virtual Bariatric Endoscopy (ViBE) simulator for training and assessment of the ESG procedure. The objectives of this current work are to: (a) perform a task analysis of ESG and (b) create metrics to be validated in the created simulator.

Methods We performed a hierarchical task analysis (HTA) by identifying the significant tasks of the ESG procedure. We created the HTA to show the breakdown and connection of the tasks of the procedure. Utilizing the HTA and input from ESG experts, performance metrics were derived for objective measurement of the ESG procedure. Three blinded video raters analyzed seven recorded ESG procedures according to the proposed performance metrics.

Results Based on the seven videos, there was a positive correlation between total task times and total performance scores ($R=0.886$, $P=0.008$). Endoscopists expert were found to be more skilled in reducing the area of the stomach compared to endoscopists novice (34.6% reduction versus 9.4% reduction, $P=0.01$). The mean novice performance score was significantly lower than the mean expert performance score (34.7 vs. 23.8, $P=0.047$). The inter-rater reliability test showed a perfect agreement among three raters for all tasks except for the suturing task. The suturing task had a significant agreement (Inter-rater Correlation = 0.84, Cronbach's alpha = 0.88). Suturing was determined to be a critical task that is positively correlated with the total score ($R=0.962$, $P=0.0005$).

Conclusion The task analysis and metrics development are critical for the development of the ViBE simulator. This preliminary assessment demonstrates that the performance metrics provide an accurate assessment of the endoscopist's performance. Further validation testing and refinement of the performance metrics are anticipated.

Keywords Endoscopic simulator · Medical education · Endoscopic sleeve gastroplasty · Hierarchical task analysis · Bariatric endoscopy

Obesity is a major global public health problem [1, 2]. For patients with obesity Class II (BMI 35 kg/m^2 to < 40 kg/m^2) or Class III (> 40 kg/m^2), the most potent treatment for effective and sustained weight loss is bariatric surgery [3]. Unfortunately, only 1% of the eligible patients receive bariatric surgical treatment for their obesity due to cost, access to care, personal preference, and associated risks [4, 5]. To fill the gap between invasive bariatric surgical procedures and intensive lifestyle and diet therapy, endoscopic sleeve gastroplasty (ESG) is rapidly emerging as a novel technique that provides an incision-less, effective alternative therapy to patients, especially for the candidates that are not otherwise eligible for surgery.

ESG is a minimally invasive non-surgical procedure that utilizes a full-thickness endoscopic suturing platform to decrease the effective size of the stomach. This procedure reduces the size of the gastric reservoir, delays gastric emptying, and induces early satiety, thereby eliciting significant weight loss [6]. The procedure is performed using

✉ Doga Demirel
ddemirel@floridapoly.edu

1 Department of Computer Science, Florida Polytechnic University, 4700 Research Way, Lakeland, FL 33805, USA

2 Department of Computer Science, University of Central Arkansas, Conway, USA

3 Department of Mechanical, Aerospace and Nuclear Engineering, Rensselear Polytechnic Institute, Troy, USA

4 Division of Gastroenterology and Hepatology, Indiana University School of Medicine, Indianapolis, USA

Published online: 29 November 2021

an endoscopic over scope suturing platform (Overstitch™, Apollo Endo-surgery, Inc, Austin, TX, USA). The ESG procedure has shown encouraging short and mid-term results. In a study conducted by Sharaiha et al. [7], a total of 216 patients (68% female) between the ages 33–59 with a BMI between 33 and 45 kg/m^2 underwent ESG. Out of the 216 patients, 38 patients were eligible for a 5-year follow up. At 5 years post-procedure, the mean total body weight loss (TBWL%) percentage was 15.9%. Furthermore, the results are shown to be durable with mid-term and long-term data accumulating [8]. With a low complication profile, a short recovery period, and suggestions of cost-effectiveness, ESG is a promising alternative treatment for patients suffering from Class 2 or Class 3 obesity.

ESG is a procedure that requires experience and training in endoscopic suturing. The procedure requires intricate endoscopic handling and multiple technical steps to successfully place full-thickness sutures in the stomach. The procedure requires the knowledge and use of multiple catheters within the system to grasp tissue, exchange the suturing needle back and forth within the tissue, and cinch sutures. Endoscopists and surgeons must be well-trained in endoscopic suturing and this specific application of endoscopic suturing to safely and effectively perform an ESG. Currently, the most common training for endoscopic suturing is from industry or medical society sponsored courses that utilize suturing in ex-vivo porcine specimens. Such specimens can be costly and are limited to one-time use. Further, the porcine anatomy is somewhat different from human anatomy, and with ex-vivo specimens, the tissue tension and lack of perfusion (i.e., no oozing or bleeding risks) lead to a lack of realism. As a reflection of the need for physician training in this realm, the American Society for Gastrointestinal Endoscopy (ASGE) provides a skill, training, assessment, and reinforcement (STAR) program for endoscopic suturing [9]. This program is a modern online curriculum that contains three parts. Those three parts are an online curriculum, a live course, and a post-course skill assessment. The hands-on portion of this program is the live course with endoscopic suturing performed on ex-vivo porcine specimens [10].

Without previous knowledge of suturing, ESG takes around thirty-five cases to attain basic proficiency [8, 11]. Endoscopists already proficient in endoscopic suturing can expect to become proficient in ESG after approximately seven actual cases [12]. To overcome such issues in training and to shorten the learning curve, virtual reality (VR)-based training may provide unique benefits to learning this procedure. Such VR tools are already being developed for endoscopy, minimally invasive [13] and open surgery [14], cricothyrotomy [15, 16], and rotator cuff [17].

Researchers have explored the effects of VR simulation on physical rehabilitation, pain management, surgery training, anatomical education, and the treatment of psychiatric disorders [18]. Compared with traditional methods, VR technology is emerging as an effective and efficient tool in the aforementioned areas [18]. Medical simulators have been shown to reduce the risk of error caused by humans and increase the safety of patients by training medical students more efficiently in a short period [19]. Using VR has improved learning in 17 (74%) studies and higher accuracy in medical practice in 20 (87%) studies [20]. VR trainers allow clinicians access to unlimited training materials that could reduce training costs and learning curve. VR based simulators have also seen significant adoption in bariatric surgery over the years [21, 22]. VR in laparoscopic surgery has been proven to be effective for training surgeons and the development of VR simulated curricula continues to grow [23, 24]. A VR simulator is now used for the Fundamentals of Endoscopic Surgery (FES), which is a pre-requisite for the Board certification in general surgery [25]. While cost has been pointed out as an obstacle in adopting VR, the rapid acceleration of computational technology and widespread adoption of VR are expected to address that issue. VR-based training simulators offer risk-free training and assessment at different levels and allow trainees to perform repetitive tasks and receive objective feedback on their performance. It is anticipated that VR training would have a positive impact on ESG and other endobariatric procedures.

Our ultimate goal is to develop and validate a Virtual Bariatric Endoscopic (ViBE) simulator that can simulate ESG procedures to train and assess trainees in this procedure. To achieve this, the first steps require an in-depth depiction of the tasks and sub-tasks associated with ESG. Hierarchical task analysis (HTA) has been shown to overcome the limitations of classical methods in analyzing complex tasks [26]. An HTA aims to distill tasks into subtasks to a level of detail in executable or observable actions [26]. Metrics derived from the HTA allow for objective evaluation of the trainees and development of goal scores for competency-based performance evaluation. We have developed and effectively used HTA for cricothyroidotomy [24], endoscopic submucosal dissection [27], and shoulder arthroscopy [28] procedures. HTAs have been developed for endoscopic sinus surgery (FESS) [29], robotic sacro colpopexy simulation [30], vertebroplasty simulator [31], medication administration [32], human factors in anesthetic practice [33], and improving patient positioning for direct lateral interbody fusion in spinal surgery [34].

In this study, we aim to create an HTA of the ESG procedure and develop a preliminary set of metrics concerning each task. Secondarily, we will use the proposed metrics to analyze the performance of expert and novice endoscopists based on a set of ESG procedural videos.

Materials and methods

The Indiana University Human Research Protection Program has determined the project does not require an IRB review due to the project not involving human subjects.

Creation of HTA: procedure analysis

The HTA was developed using an iterative process, led by two experts in advanced endoscopy with clinical experience in ESG. The process included an extensive review of the medical literature regarding the technique of ESG, analysis of ESG procedural videos, and experts' opinions. The ESG procedure was divided into tasks and subtasks, with optional tasks noted when applicable. The tasks and subtasks were then organized into a hierarchical task tree based on the expected chronological order of tasks and subtasks.

Creation of HTA: identification of operative errors

Based on the literature review and expert review, we compiled the most significant errors associated with ESG in Table 1. During the diagnostic Esophagogastroduodenoscopy (EGD) phase, the primary error that (rarely) occurs is oropharyngeal trauma and perforation of the examined parts of the upper gut including esophagus, stomach, and proximal duodenum. The resulting complications can be a free flow of air, fluid leak, and infection. As for the suturing phase, multiple errors can occur. Suturing errors may lead to incomplete gastric plication, perforation, bleeding, premature displacement of suture, inadequate or excess suture tension, and incorrect suture pattern.

Metric creation

Currently, no standardized quality or procedure metrics exist for ESG. Thus, there is no existing set of metrics that have been devised or validated. Utilizing the information from the hierarchical task analysis and with input from endoscopic experts in ESG, performance metrics were derived for objective measurement of the ESG procedure. The hierarchical tasks and subtasks were analyzed on an individual basis. When a desired/less desired/undesired option existed for a task/sub task, this was incorporated into the devised metrics.

The metric scoring system incorporates performance components that account for optimal and suboptimal actions and the time to complete the procedure and its tasks, similar to prior performance metric systems for virtual reality simulators [27]. To most effectively incorporate time as a function of the metrics, our proposed system utilized an inverted graded scoring system with the best action on a task given as a score of 0, an acceptable action a score of 3, and a failed attempt a score of 5, where applicable. Therefore, a lower

Table 1 Operative errors and complications

Phase	Errors and mistakes	Occurrence and complications
Diagnostic EGD	• Oropharyngeal trauma • Perforation [35]	• Free flow air • Fluid leakage • Infection
Suturing [36]	• Perforation • Bent suturing needles • Loose suture (incorrect cinch) • Not enough suture bites •Non-full thickness endoscopic sutures • Prematurely released suture • Bleeding/ Oozing blood • Esophageal Oropharyngeal tear/trauma • Incorrect suture location • Damage to adjacent organs	• Bleeding [37] • Excessive tension • Blood clot forming • Incomplete or incorrect gastric plication • Abcess formation • Extragastric bleeding or organ damage • Tear or perforation
Overall procedure/general surgical risks	• Prolonged anesthesia (due to long completion time) [38]	• Thromboembolism • Post-operative infection • Hypothermia • Fever with no procedure-related collection • Perigastric collection with bilateral pleural effusion • Perigastric collection with left-sided pleural effusion • Severe abdominal pain/nausea •Readmissions + conservation management • Readmission + reversal of ESG [39]

overall score signifies better performance. Since we are also dealing with time metrics where longer time means a worse performance, inverting our metrics was the best decision.

Analysis of ESG videos with relation to tasks and metrics

We analyzed seven full-length ESG procedure videos performed on ex-vivo porcine specimens for a separate training study. There were videos from endoscopists both with and without significant endoscopic suturing experience. Endoscopists considered an expert in endoscopic suturing performed four procedures (2 experts performed 2 procedures each), and endoscopists considered novice to endoscopic suturing performed three procedures (3 novices performed one procedure each). The expert determination was based on their previous experience performing clinical ESG for clinical indications and at least 50 cases with the endoscopic suturing device. Novice determination was based on no prior clinical ESG experience and < 5 independent cases with the endoscopic suturing device. All participants had at least 4 years of clinical endoscopy experience, with at least 1 year of advanced endoscopy training. Each ESG procedure was recorded from 3 different views using (1) a video camera mounted on the endoscopist's head; (2) the endoscopic view; (3) a stationary video camera focused on the endoscope and endoscopist's hands. Representative views are illustrated in Fig. 1. During the ESG procedures, an expert in ESG was available for those in the novice group to answer procedural or technical questions, if prompted.

We analyzed the seven procedural videos to assess whether the HTA captured all the essential steps and their variations (e.g., differences among endoscopists). In our timing analysis, we analyzed each procedural video and measured the precise task and procedure times using Windows and VLC media players. For each major task that was defined in our HTA, we identified the start and end times corresponding to each task (Table 2). A group of three external raters (not involved in the endoscopic procedures) was used to determine the timing and scores for each video based on the procedure-specific metrics created. The total time for each task as well as the time to complete the entire procedure was collected for each performance and used for descriptive statistical analysis.

Before analyzing our data, we hypothesized that the time it takes to perform certain tasks, as well as the entire procedure, would be closely correlated to the task scores and the total scores. This hypothesis is supported by previous studies where the experience and skill level of surgeons affected the total duration of the surgery and specific tasks [24, 28]. We performed Pearson's correlation tests between times and scores by using R studio alongside Microsoft Excel for basic data manipulation.

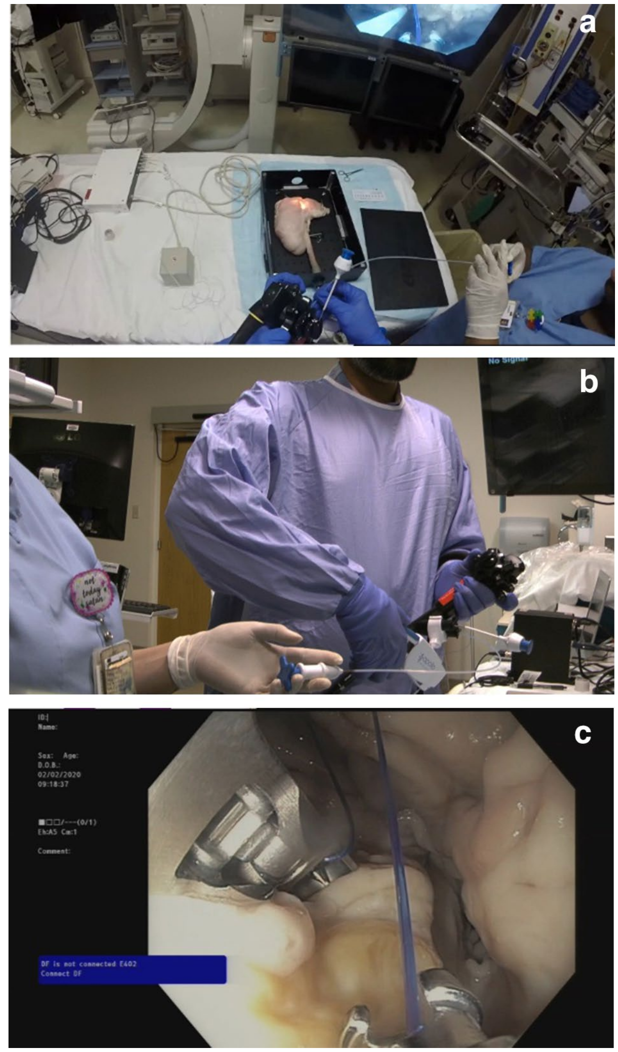

Fig. 1 ESG procedures were recorded from three different views, **a** head-mounted camera view, **b** stationary side view, and **c** endoscope view

Results

Hierarchical task analysis

The hierarchical task analysis (HTA) is organized according to a hierarchy of phases, tasks, and subtasks. Each phase is divided into tasks and, where appropriate, sub-tasks. In our HTA, the tasks are categorized under three main phases: (a) pre-procedure; (b) evaluation; and (c) ESG procedure. The HTA and all three phases and their tasks are shown in Fig. 2. In the graphical representation of this HTA, arrows in figures indicate the linear progression of the procedure, and each branch under the phases and tasks represents the progression of task execution. The four main tasks in the HTA that make up the three phases are preparation, checking tools, diagnostic Esophagogastroduodenoscopy (EGD), and suturing.

Table 2 Start and end time events of each main task

Task	Start time event	End time event
Preparation	General anesthesia	Pre-procedure DVT prophylaxis finished
Check tools	Start gather tools	All tools gathered
diagnostic EGD	Insertion of single channel Diagnostic gastroscope started	APC catheter inserted into the working channel of EGD scope (if optional marking is chosen then marking the greater curvature will be the end event for this task)
Suturing	Mounting of overstitch suturing platform on double-channel therapeutic gastroscope started	Double channel gastroscope removed
Total procedure	Start of procedure	End of procedure

Based on expert opinion and review of the available literature, it was appreciated that there is some variability in the technical aspects of ESG. Thus, where applicable, optional or variable tasks were outlined where able.

The first phase in HTA is the preparation phase, which begins with sedation (often general anesthesia followed by endotracheal intubation). Although not common, some centers do perform ESG with moderate or deep conscious sedation without endotracheal intubation. After anesthesia, the patient is placed in the left lateral decubitus position, given prophylactic antibiotics and deep venous thrombosis prophylaxis. The endoscopist assesses that all instruments are complete and prepared for the procedure. After the endoscopist confirms that all tools are present, they will advance to the diagnostic EGD phase that commences by inserting the single-channel diagnostic gastroscope (and the overtube, if used [40]) in the standard fashion. The next step in the diagnostic EGD phase is for the endoscopist to confirm the absence of any exclusion criterion for ESG, including presence of prior gastric surgery, gastric ulcerations, gastric varices, portal hypertensive gastropathy, or lesion suspicious for dysplasia or malignancy. The endoscopist will then insert the APC catheter into the working channel of the EGD scope. This phase concludes with the marking with interrupted lines of the three main points in the stomach—the anterior, posterior, and the greater curvature. The marking phase is identified as an optional task for experts in ESG; therefore, this step is indicated as optional by a yellow background on each subtask in the HTA representation. If the esophageal overtube is used (optional task), then at this time the overtube may be placed, prior to the suturing phase.

The last phase is the suturing phase, which begins by mounting the overstitch suturing platform on the double channel therapeutic gastroscope (the optional task is the use of the single-channel suturing system on a single channel therapeutic gastroscope). There is an alternative, newer suturing system (Apollo Endosurgery) designed for single channel therapeutic gastroscope [41]. The endoscopist must ensure that the suturing arm is closed and then load the first suture onto the suturing arm. They will then delicately insert the double-channel gastroscope into the oropharynx and then advance into the esophagus down into the stomach. Once the endoscopist has entered the stomach, they are required to identify the location of the first bite. The next step is to grasp the target tissue with the tissue helix. When grasping target tissue and using the cinch device, there are multiple subtasks involved, indicated with "**" and "~" respectively, in Fig. 2.

When grasping the target tissue with the tissue helix, the endoscopist will first advance the catheter within the scope channel and choose the targeted tissue to grasp. The endoscopist will then instruct the assistant endoscopist to extend the helix device and rotate the blue cross 3–4 turns clockwise. The endoscopist will then pull the tissue with the catheter within the arm of the suturing device. Once this is completed and the endoscope is maneuvered to achieve optimal tissue capture, the endoscopist will close the suturing arm, transfer the suture to the catheter, reopen the arm, advance the tissue helix catheter with the endoscopist directing the assistant to rotate the blue cross 3–4 turns counterclockwise to release the previously captured tissue and then retract the helix device. The steps of suturing bites are repeated in a running fashion (i.e., the anterior wall, the greater curvature, and the posterior wall) for a series of sutures, until the entire gastric lumen is sutured, up to the fundus. The number of bites per suture, the number of sutures per procedure, and the suture pattern are variable, but could affect the impact on the safety and effectiveness of the procedure. There are ongoing studies to help determine differences in suture placement and pattern. Despite their significance, the optimal use of sutures and pattern are still of high salience to the endoscopists [42]. It is reported that using fewer sutures could reduce procedure cost and time, which could also simplify ESG training [43]. At the moment, a common practice for ESG is about 6–7 sutures and 6–11 bites per suture. Different patterns can be used,

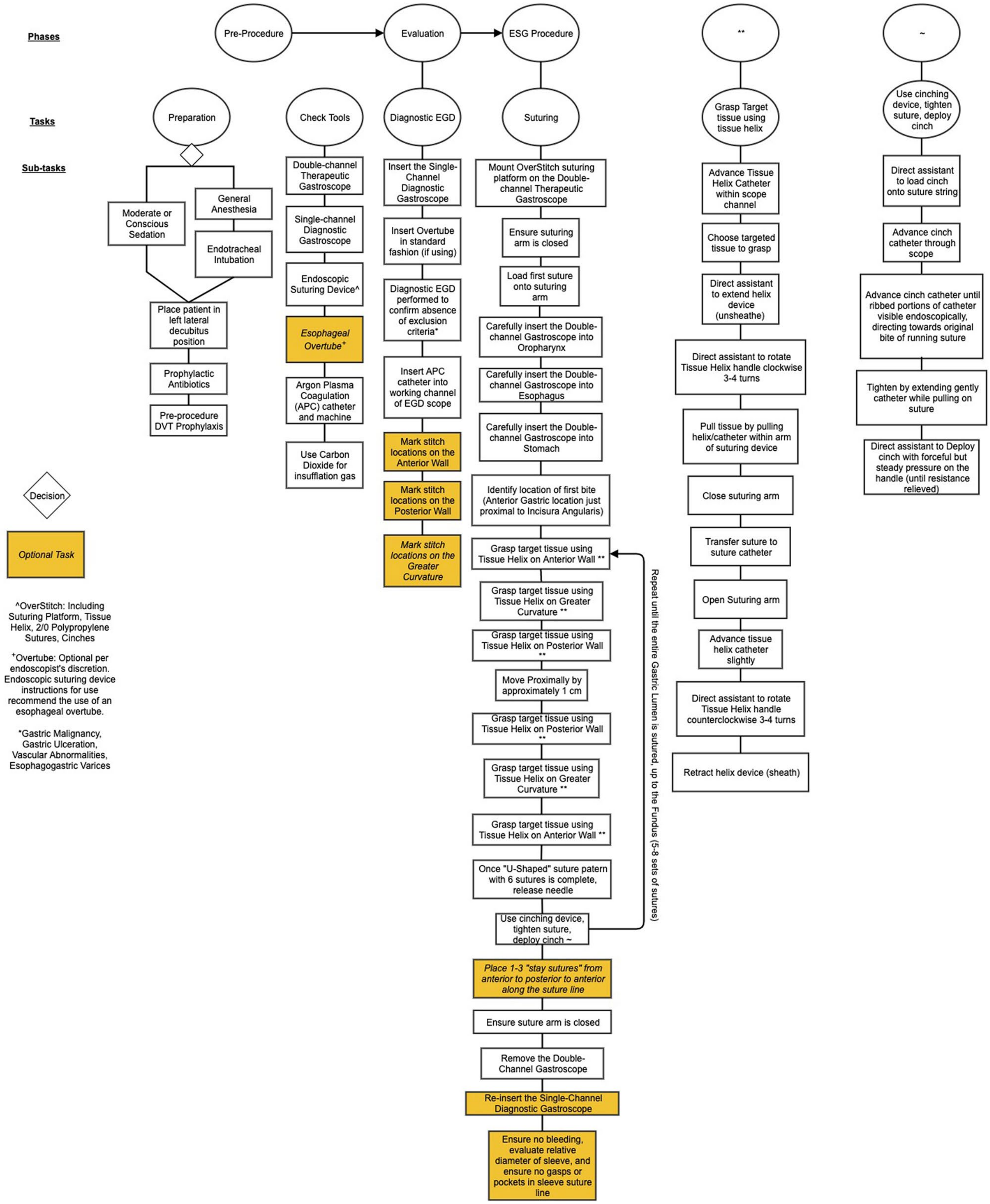

Fig. 2 Hierarchical decomposition tree for ESG phases. The forward arrows indicate a linear progression of the procedure. Sub-tasks involved in the tissue grasping and cinching are shown with "**" and "~", respectively

however, the most common one is the "modified U pattern" which goes from the anterior wall to the greater curvature, then to the posterior wall, back to the greater curvature, and back to the anterior wall.

The cinch device is used to tighten the sutures. Each time the cinch device is used, the endoscopist begins by directing the assistant to load the cinch onto the suture and advancing the cinch catheter through the scope until the ribbed portions of the catheter are visible endoscopically. The endoscopist then tightens the suture by extending the catheter gently while pulling on the suture and asking the assistant to deploy the cinch by applying pressure on the handle. This step is repeated just as the grasping subtask is repeated for each suture.

After completion of all sutures on the initial sleeve, a set of sutures called reinforcement sutures, are placed approximating the anterior and posterior walls of the plicated tissue of the sleeve to 'reinforce' the primary sutures. The use of reinforcement sutures may relatively improve weight reduction in one-year follow-up [8]. The created gastric sleeve is then examined with the endoscope, ensuring adequate sleeve creation and the presence of no complications such as perforation or ongoing bleeding. The last necessary steps of the suturing phase are to ensure that the suturing arm is closed, which is then followed by the removal of the double-channel gastroscope. If used, the overtube is then removed.

Scoring metrics

The scoring system incorporates performance components that account for optimal and suboptimal actions and the time to complete the procedure and its tasks, similar to prior performance metric systems for virtual reality simulators [27]. The details of the metrics are shown in Figs. 3, 4, 5, 6, 7, 8, 9, and 10. Our ESG metrics utilized an inverted scoring system with the best action on a task given as a score of 0, an acceptable action a score of 3, and a failed attempt a score of 5. Therefore, a lower overall score signifies better performance, and the perfect score is zero.

Based on the HTA, we assigned metrics for seven different main tasks and each of their subtasks. Performance metric criteria were also created for complications and for the task times of the procedure. The overall time of the procedure was based on quartiles created from ex-vivo procedural videos.

In ESG, for both EGD and APC, the tools are inserted into the pharynx, then into the esophagus, and finally into the stomach. Each of these steps is assigned a score of 0 if there are no perforations and a score of 5 if there are perforations. The next step is marking, which is optional but will be scored if it is performed. Insertion of EGD and APC and optional marking metrics are seen in Figs. 3 and 4, respectively.

Fig. 3 Insertion of EGD and APC metrics

No	Metrics	Score
	Insertion and Diagnostic Upper Endoscopy	
1.	**Insert Over tube in standard fashion (Optional)**	
	No perforation	0
	Perforation	5
2.	**Endoscope inserted into posterior pharynx**	
	No perforation	0
	Perforation	5
3.	**Advance endoscope into esophagus**	
	No perforation	0
	Perforation	5
4.	**Diagnostic Evaluation of Esophagus**	
	Diagnostic Evaluation performed	0
	Not performed	5
5.	**Advance endoscope into stomach**	
	No perforation	0
	Perforation	5
6.	**Advance endoscope into duodenum**	
	No perforation	0
	Perforation	5
7.	**Diagnostic Evaluation of duodenum**	
	Diagnostic Evaluation performed	0
	Not performed	5
8.	**Diagnostic Evaluation of Stomach**	
	Diagnostic Evaluation performed	0
	Not performed	5

Fig. 4 Optional marking metrics

APC Marking (Optional)		
9.	**Advance argon plasma coagulation tool through channel**	
	Tool advanced appropriate amount	0
	Tool was not advanced appropriate amount	5
10.	**Mark Anterior Wall**	
	Parallel line	0
	Non- parallel line	3
	No mark	5
11.	**Mark Posterior Wall**	
	Parallel line	0
	Non- parallel line	3
	No mark	5
12.	**Mark Greater Curvature**	
	Parallel line	0
	Non- parallel line	3
	No mark	5

Fig. 5 Insertion of the suturing arm metrics

Insertion of suturing arm		
13.	**Mount overstitch suturing platform on double channel therapeutic gastroscope**	
	Mounted/suturing arm closed/first suture loaded	0
	Not mounted/not closed/ not loaded	5
14.	**Insert Double channel Gastroscope into pharynx**	
	No perforation	0
	Perforation	5
15.	**Insert Double channel Gastroscope into Esophagus**	
	No perforation	0
	Perforation	5
16.	**Insert Double channel Gastroscope into stomach**	
	No perforation	0
	Perforation	5

After the optional marking step, the next task is to insert the suturing arm, as shown in Fig. 5 below. We score the overstitch mounting with a 0 if it is properly installed, the suturing arm is closed, and the first suture is loaded, but if any of these sub-steps is not performed, the endoscopist will earn a score of 5. The endoscopist will be given a 0 for adequately inserting the therapeutic gastroscope into the pharynx, 0 for inserting the therapeutic gastroscope into the esophagus, and a 0 for inserting the therapeutic gastroscope into the stomach.

The next major step in our metrics is to evaluate the suturing process shown in Figs. 6 and 7 below. The endoscopist will be given a 0 for starting the suturing at or proximal to the incisura angularis and a score of 0 for grasping the tissue in the proper area. Then they will be given a 0 for suturing near the APC mark (if used) with a parallel line or a 3 for a non-parallel line. This step is then repeated for the anterior wall, greater curvature, and the posterior wall. They will be graded for the suture order, suture direction, bite amount per suture, the amount of suture, whether the suture is tightened, the suture line, and the location of the last suture.

We developed metrics for possible complications. The scoring criteria are associated with endoscopists tackling specific issues such as severe bleeding or bent sutures (See Fig. 8). The task time is an important measure in discriminating the experts and novices [27]. We measure the time performance of each main task and sub-tasks of the endoscopists. This can be seen in Fig. 9. There are four quartiles, each measured by seconds, for the first quartile, the total duration of the procedure must be under 2789 s; the second quartile is 280–4484 s, the third quartile is 4485–6254 s, and the fourth quartile is 6255 s or more. The quartiles are calculated as; Quartile 1- is the lowest 25% of the time values, Quartile 2-time values that are between 25% to the median, Quartile-3 is 25% above the median, and Quartile-4 is the highest 25% of the time values. The last metric is to score the communication skills between the endoscopist and their assistant technician, which can be seen in Fig. 10. Of note, due to the nature of the time and communication metrics, the scoring paradigm was changed slightly (instead of previously used ideal score 0, suboptimal score 5).

Fig. 6 Suturing metrics

Suturing		
17.	**Start of suture**	
	Start just proximal to incisura angularis on anterior gastric wall	0
	Start at different location	5
18.	**Grasp Tissue on anterior wall**	
	Grasp near marked tissue (within 0.5 cm of marking)	0
	Grasp away from than marked tissue	5
19.	**Suture Anterior Wall**	
	Correctly complete suture and anchor exchange	0
	Incorrectly complete suture and anchor exchange	5
20.	**Grasp Tissue on greater curvature**	
	Grasp marked tissue (within 0.5 cm of marking)	0
	Grasp away from marked tissue	5
21.	**Suture Greater Curvature**	
	Correctly complete suture and anchor exchange	0
	Incorrectly complete suture and anchor exchange	5
22.	**Grasp Tissue on posterior wall**	
	Grasp marked tissue (within 0.5 cm of marking)	0
	Grasp away from marked tissue	5
23.	**Suture Posterior Wall**	
	Correctly complete suture and anchor exchange	0
	Incorrectly complete suture and anchor exchange	5
24.	**Suture Direction (after each anterior/greater curve/posterior suture series)**	
	Distal to Proximal 1-2 cm	0
	Any other Direction/Amount	5
25.	**Suture Bite (per bite)**	
	Full thickness	0
	Any other bite	5
26.	**Number of Bites per running suture**	
	6 to 7 bites	0
	More than 7	3
	Less than 6	5
27.	**U-Shaped pattern (if using u shaped pattern)**	
	yes	0
	no	5
28.	**Tighten Sutures**	
	Release T tag correctly to form a plication using cinching device	0
	Don't release T tag correctly (or premature deployment of T-tag)	5

We hypothesized that our metrics would distinguish between skill levels with experts attaining lower average scores (better score) than the novices. After the seven procedure videos were scored using the above derived metrics, we performed the Shapiro Wilk test on the total performance scores and had a $P = 0.979$, indicating normality in the data distribution. The average novice total score was 34.67, which differed by 10.92 points from the average expert score of 23.75, indicating a significant difference ($P = 0.047$). The overall scores for novice and expert endoscopists can be seen in Fig. 11.

As time is a possible key indicator of performance, as noted in the literature [8, 9], Pearson's correlation test was performed to find any correlations between the task times, expertise level, and the performance scores. We observed a strong positive correlation between the total scores and the total time for completion ($R = 0.893$, $P = 0.007$). Figure 12 shows a scatter plot of the total score vs the total duration. We also observed a strong correlation of the total score with suturing performance ($R = 0.962$, $P = 0.0005$). Figure 13 below contains two box plots that illustrate the average time completion score for expert (2.25) and novice (6.0) participants.

The suturing task was the task with the most variance in scores. Pearson's correlation test showed that the number of sutures performed had a strong positive correlation ($R = 0.882$, $P = 0.009$) with experience level, while suturing and its sub-task, suture bite thickness, also showed a strong positive correlation ($R = 0.895$, $P = 0.006$). Figure 14 below shows two box plots that illustrate the comparison of average

Fig. 7 Suturing metrics continued

29.	**Suture Line**	
	Bring suture line proximally within 1-2 cm of gastroesophageal junction on the lesser curvature	0
	Bring suture line proximally either < 1cm from GE junction or 2-3 cm from gastroesophageal junction on the lesser curvature	3
	Bring suture line proximally more than 3 cm of gastroesophageal junction on the lesser curvature	5
30.	**End suture**	
	Do not suture fundus	0
	Suture within fundus	5
31.	**Evaluate sleeve for 'need for reinforcement sutures'**	
	yes	0
	no	5
32.	**Deploy reinforcement sutures *if needed***	
	Do not suture fundus	0
	Suture within fundus	5
33.	**Suture Set**	
	5-8 sets of sutures	0
	More than 8	3
	Less than 5	5
34.	**Removal of double channel gastroscope**	
	No perforation	0
	Perforation	5

Fig. 8 Complications metrics

	Complications	
35.	**Severe Bleeding**	
	Premature cinch (to stop bleeding) within 60 seconds	0
	No premature cinch within 60 seconds	5
36.	**Bent Transfer Tag**	
	Premature cinch (to prevent damage to scope)	0
	No premature cinch	5

Fig. 9 Time completion metrics

	Time completion	
37.	**Total time**	
	First Quartile	0
	Second Quartile	3
	Third Quartile	6
	Fourth Quartile	9

suturing scores for expert and novice participants. The remaining significant correlations are presented in Table 3.

Figure 15 below compares expert and novice for the time it took to perform the first stay suture. The box plot shows that the novices took longer to complete the first stay reinforcement suture. We ran the Shapiro Wilk test on stay suture time and had a P value of 0.393, indicating normality in the data distribution. A Wilcoxon test showed marginally significant evidence of a difference between experts and novices ($P = 0.057$).

Figure 16 below shows the expert and novice times for mounting the overtube and loading the suture times. The box plots show that the experts took significantly less time than the novice did. We ran the Shapiro Wilk test on the mounting time ($P = 0.023$) and loading time ($P = 0.0745$) showed that the mounting time was not normally distributed while loading time is normally distributed. A two-sample t test assuming unequal variances showed a significant difference between the expert and novice groups for the loading time ($P = 0.0013$).

Fig. 10 Communication metrics

Communication		
38.	Load Tissue Helix command (repeat for each suture set)	
	Yes	0
	No	1
39.	Extend Tissue Helix command (repeat for each bite)	
	Yes	0
	No	1
40.	Rotate blue cross on Tissue Helix clockwise 3-4 times command (repeat for each bite)	
	Yes	0
	No	1
41.	Rotate blue cross on Tissue Helix counterclockwise 3-4 times command (repeat for each bite)	
	Yes	0
	No	1
42.	Retract Tissue Helix command (repeat for each bite)	
	Yes	0
	No	1
43.	Remove Tissue Helix command (repeat for each suture set)	
	Yes	0
	No	1
44.	Load Cinch onto suture string command (repeat for each suture set)	
	Yes	0
	No	1
45.	Deploy Cinch command (repeat for each suture set)	
	Yes	0
	No	1

Operative times and inter-rater reliability

All three raters timed and graded each procedure using the metrics derived from the HTA. IBM SPSS 26 software was used to compute the inter-rater class correlation coefficient (ICC) and Cronbach's alpha. The three raters recorded the times for all seven of the procedures and each corresponding task based on each task's start and end time. The inter-rater reliability test showed a perfect agreement (ICC = 1.0, Cronbach's alpha = 1.0) among all the raters for all tasks except for the suturing task. The suturing task had a significant agreement (ICC = 0.84, Cronbach's alpha = 0.88) between the raters. Three out of the four experts operated faster than all the novice endoscopists. The slowest expert endoscopist operated in 4594 s while the fastest novice operated in 4484 s. The distribution can be seen in Fig. 17.

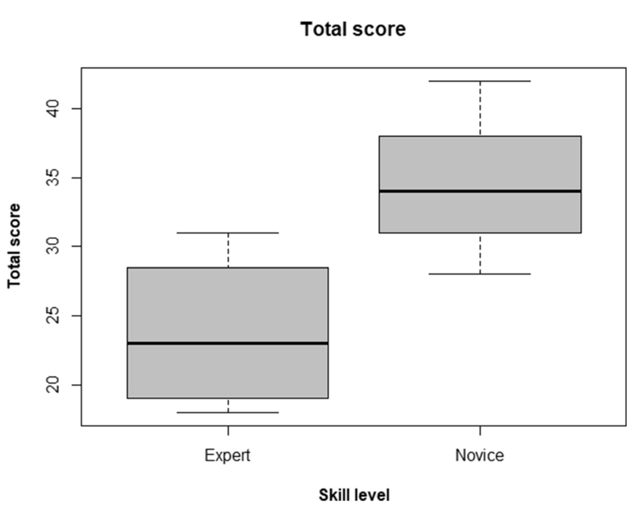

Fig. 11 The average total score of experts and novices

Post-procedure stomach analysis

We performed image processing to quantify the percentage in the reduction of the stomach area after ESG in the seven ex vivo procedures analyzed. Images of the specimens before and after the procedure were taken using an iPhone 11 with dual 12MP supporting ultra-wide aperture (ƒ/2.4) and 120° field of view. For reference and calibration, a marked ruler was placed in the experiment field near the specimen. Using ImageJ [44] analyzer (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA), areas of all the specimens were calculated by a single reviewer (see Fig. 18).

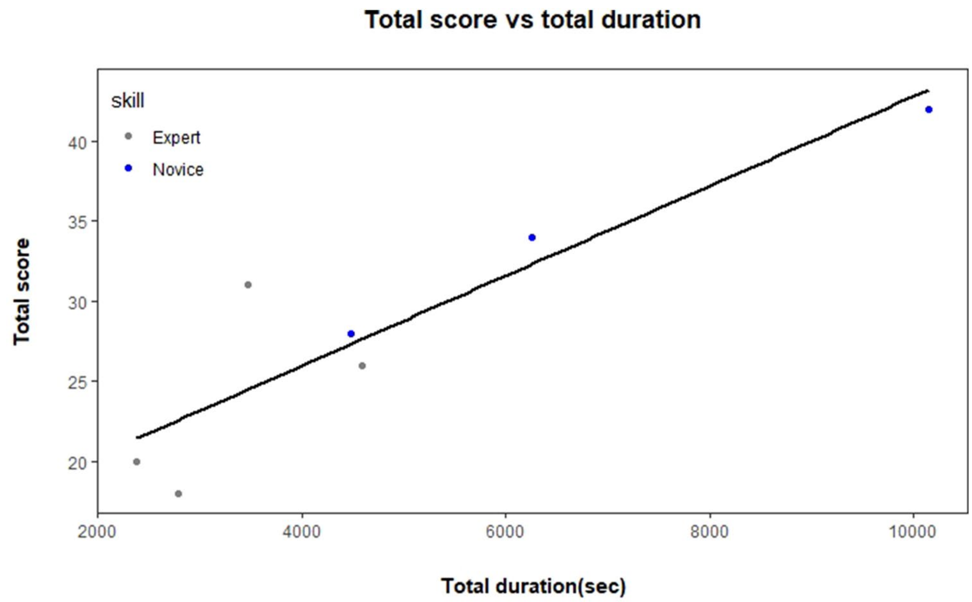

Fig. 12 Total score vs total duration scatter plot

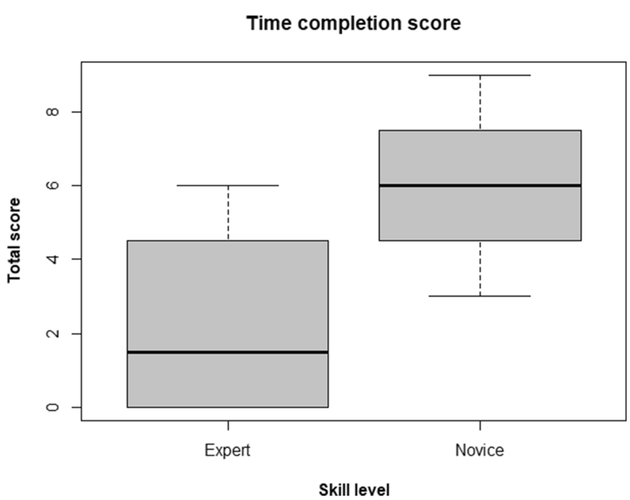

Fig. 13 Average time completion scores by skill level

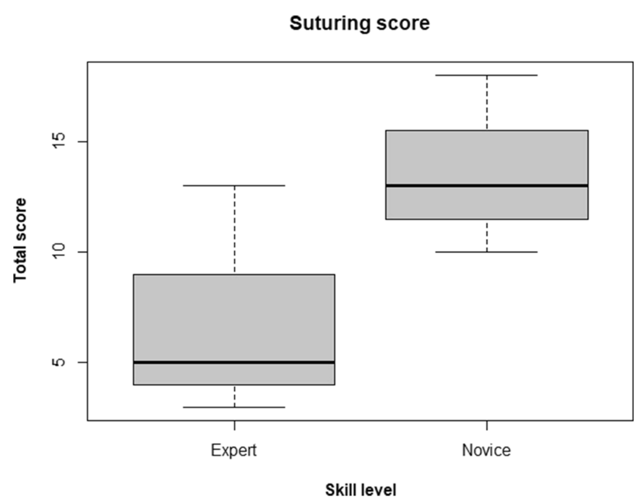

Fig. 14 Average suturing scores by skill level

Based on our results, novice endoscopists reduced the stomach an average of 9.38% compared to 34.62% by experts, as seen in Fig. 19. We ran the Shapiro Wilk test on stomach reduction and had a *P* value of 0.485, indicating normality in the data distribution. A two-sample *t* test assuming unequal variances showed a significant difference between experts and novices ($P=0.01$).

Discussion

In this paper, we presented a hierarchical task analysis of the ESG procedure, developed metrics of performance evaluation for the ESG, and analyzed the performance of expert and novice endoscopists based on ESG procedural videos. These are all important steps in the development of our high-fidelity, virtual reality ESG simulator. The HTA demonstrates that the ESG is a procedure with a complex set of tasks and sub-tasks, which should be learned and practiced by trainees.

Based on our scoring metrics, a positive correlation between time and the total score was observed. Furthermore, there was a significant correlation between experience with prior suturing and both total time and total score. Both these correlations suggest that the proposed metrics are deemed useful to distinguish between prior experience and skill level.

Table 3 Significant Pearson's correlation test results

Correlation task/s and time/s	Pearson's correlation (*R*)	Significance (*P*)
Change in area of stomach and mounting the over tube time	−0.925	0.003
Change in area of stomach and loading first suture time	−0.949	0.001
Mounting time and loading time	0.871	0.011
Total procedure time and change in the area of the stomach	−0.823	0.023
First stay suture and mounting time	0.904	0.005
First stay suture and loading time	0.827	0.0219
Total time for completion and evaluation of sleeve for 'need for reinforcement sutures'	0.881	0.009

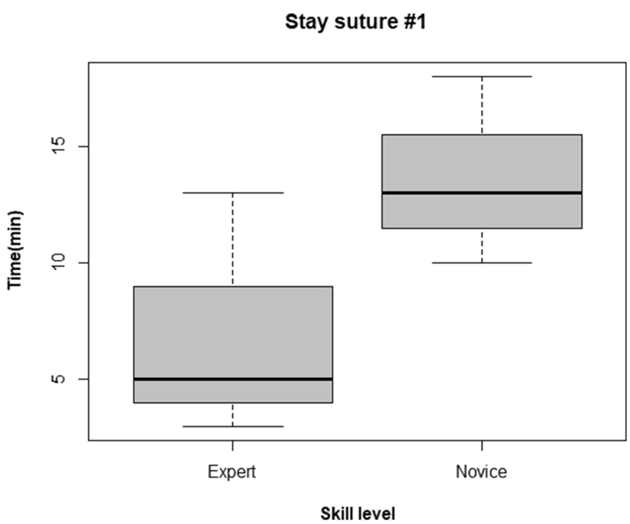

Fig. 15 Time for first stay suture

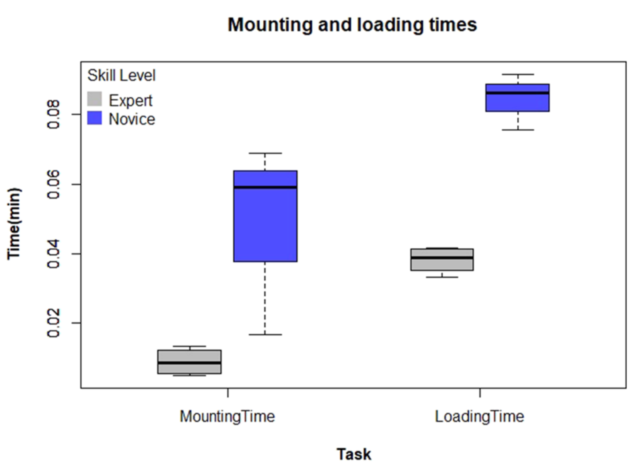

Fig. 16 Mounting and loading time by skill level

A strong positive correlation between the total time for completion and the evaluation of the sleeve for 'need for reinforcement sutures' indicates that the endoscopists that did not effectively perform this evaluation are more likely

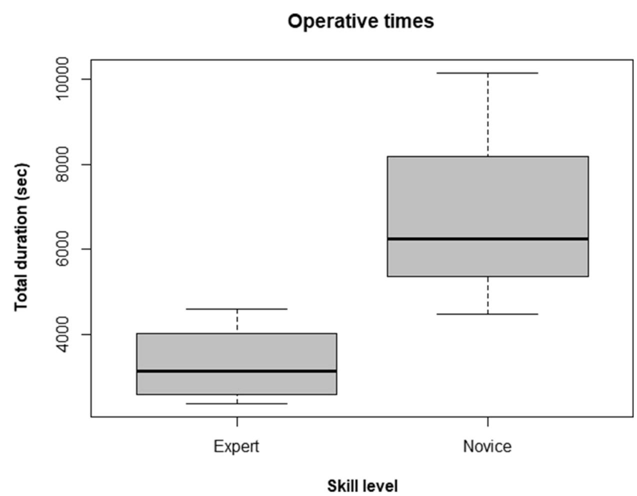

Fig. 17 Operative times

to be novices. Likewise, a strong negative correlation was observed between the total procedural time and stomach area reduction, therefore as the time it took to perform the entire procedure increased, stomach area reduction decreased. Our reasoning and also an expected observation is that expert endoscopists took less time and they were also more effective in reducing the stomach area.

We observed a strong positive correlation with the total score and the task of suturing, whether each bite was a full-thickness bite, and the total time for completion. This shows that when the total score was higher (i.e., lower performance), so was the suturing scores, specifically suture bite thickness, and the total time it took to complete the procedure. Since suture bite thickness is a subtask of the main suturing task, there was an expected strong positive correlation between the two. Since suturing was the only task that correlated with the total score, it is the most critical part of the procedure.

There are some limitations to this study. First, we had a limited number of procedural observations ($n = 7$). There is some degree of variability in the clinical and described technique of the ESG between endoscopists. This is

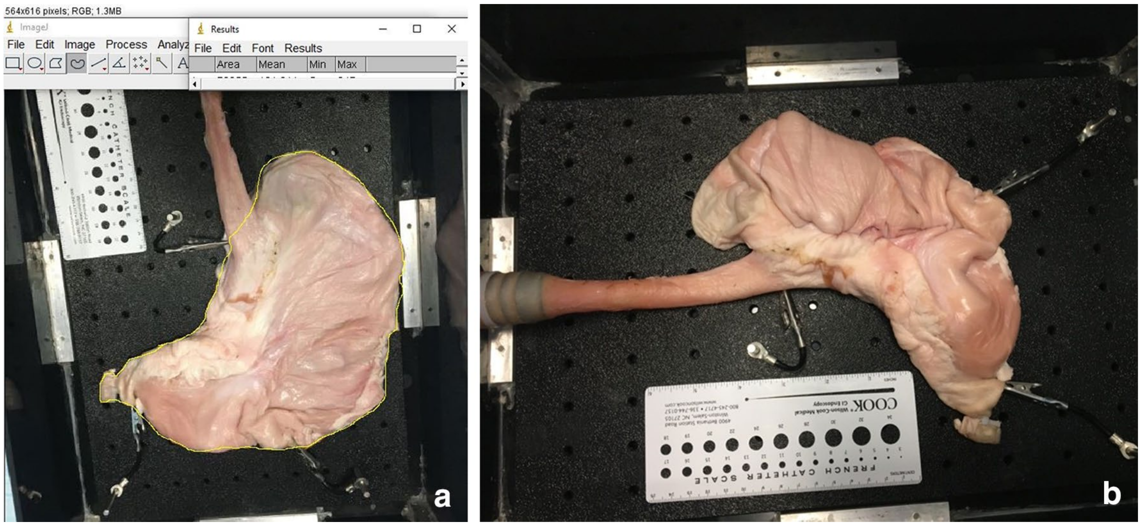

Fig. 18 **a** Porcine stomach specimen before and **b** right after the ESG procedure. The yellow highlighted region in (**a**) is an area estimation using ImageJ analyzer (Color figure online)

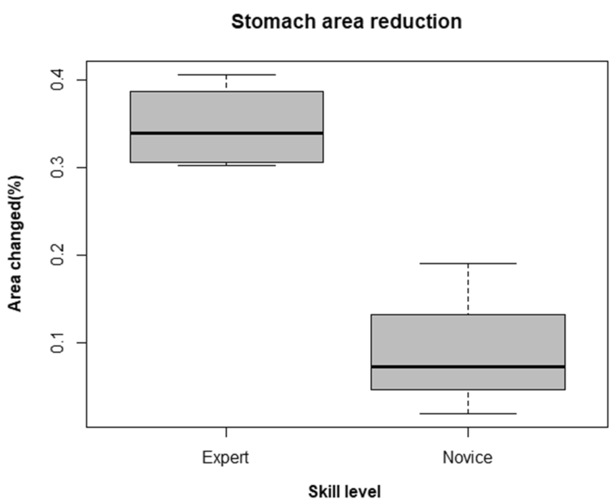

Fig. 19 The average reduction of the stomach by skill level

expected with an endoscopic or surgical procedure but makes the creation of a precise HTA somewhat difficult. Where variability was identified, we made denotations of a sub-task being optional. The devised HTA is not meant to be a standard or expected procedural set of instructions, but instead a most commonly accepted composition of the procedural parts. With regard to the procedural metrics, there are no prior metrics within bariatric endoscopy or endoscopic suturing to guide the creation of the above metrics. Thus, future studies will be expected to validate the metrics, and future modification and refinement of the proposed metrics could be anticipated.

Conclusion

With the long-term goal of improving training and performance assessment for ESG using simulator-based training, we presented an HTA based on expert knowledge of the procedure. Although the analysis presented here will pave the way to develop the clinically relevant and authentic high-fidelity simulator, it can also be utilized for assessment and feedback in ESG training using a conventional setting. In our study, we analyzed videos of both experts and novice endoscopists who performed the ESG procedure on ex-vivo porcine specimens. We found correlations for all tasks and subtasks in our graded results and the specific times it took to perform those tasks. When analyzing these correlations, we searched for defining characteristics of both experts and novice endoscopists. One of the informative conclusions from Pearson's correlation test results showed the number of sutures performed had a strong positive correlation ($R = 0.882$) with years of experience.

Not only did we identify the important tasks of the ESG procedure, but we also found that the total duration of the procedure can indicate the level of expertise of an endoscopist. This can be seen when looking at the correlation between the total duration and area changed ($R = -0.823$). The longer th e procedure took, the less effective the endoscopist was at reducing the size of the stomach-a likely indication of lack of experience. We would like to note that the sample size in our analysis is a limitation.

We plan to incorporate the findings of our HTA and procedural metrics into the creation of our VR-based ESG simulator. Further validation of these will be performed in advanced phases of the simulation development.

Funding This project was made possible by the Arkansas INBRE program, supported by a grant from the National Institute of General Medical Sciences, (NIGMS), P20 GM103429 from the National Institutes of Health (NIH). This project was also supported by NIH/NIAMS R44AR075481-01, NIH/NCI 5R01CA197491, and NIH/NHLBI NIH/NIBIB 1R01EB025241, R56EB026490.

Declarations

Disclosures James Dials, Drs. Doga Demirel, Tansel Halic, Suvranu De, Adam Ryason, Shanker Kundumadam, Mohammad Al-Haddad, and Mark Gromski, have no conflicts of interest or financial ties to disclose.

References

1. Ogden CL, Carroll MD, Kit BK, Flegal KM (2014) Prevalence of childhood and adult obesity in the United States, 2011–2012. JAMA 311(8):806–814. https://doi.org/10.1001/jama.2014.732
2. M. dos P. Galvão-Neto, E. Grecco, T. F. de Souza, L. G. de Quadros, L. B. Silva, and J. M. Campos (2016) "Endoscopic sleeve gastroplasty—minimally invasive therapy for primary obesity treatment," *ABCD, arq. bras. cir. dig* 29(1): 95–97. https://doi.org/10.1590/0102-6720201600s10023.
3. Kissler HJ, Settmacher U (2013) Bariatric surgery to treat obesity. Semin Nephrol 33(1):75–89. https://doi.org/10.1016/j.semnephrol.2012.12.004
4. Mechanick JI et al (2013) Clinical practice guidelines for the perioperative nutritional, metabolic, and nonsurgical support of the bariatric surgery patient—2013 update: cosponsored by American association of clinical endocrinologists, the obesity society, and American society for metabolic & bariatric surgery. Surg Obes Relat Dis 9(2):159–191. https://doi.org/10.1016/j.soard.2012.12.010
5. Birkmeyer JD 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
6. Abu Dayyeh BK et al (2017) Endoscopic sleeve gastroplasty alters gastric physiology and induces loss of body weight in obese individuals. Clin Gastroenterol Hepatol 15(1):37-43.e1. https://doi.org/10.1016/j.cgh.2015.12.030
7. Sharaiha RZ et al (2020) Five-year outcomes of endoscopic sleeve gastroplasty for the treatment of obesity. Clin Gastroenterol Hepatol. https://doi.org/10.1016/j.cgh.2020.09.055
8. Sharaiha RZ et al (2017) Endoscopic sleeve gastroplasty significantly reduces body mass index and metabolic complications in obese patients. Clin Gastroenterol Hepatol 15(4):504–510. https://doi.org/10.1016/j.cgh.2016.12.012
9. "STAR Certificate Programs," *Default*. https://www.asge.org/home/education/advanced-education-training/star-certificate-programs. Accessed 16 Dec 2020.
10. Bazarbashi AN (2020) Training in bariatric endoscopy. ACG Case Rep J. https://doi.org/10.14309/crj.0000000000000358
11. Lopez-Nava G et al (2017) Endoscopic sleeve gastroplasty for obesity: a multicenter study of 248 patients with 24 months follow-up. Obes Surg 27(10):2649–2655. https://doi.org/10.1007/s11695-017-2693-7
12. Hill C et al (2017) Endoscopic sleeve gastroplasty: the learning curve. Endosc Int Open 5(9):E900–E904. https://doi.org/10.1055/s-0043-115387
13. Farmer J et al (2020) Systematic approach for content and construct validation: case studies for arthroscopy and laparoscopy. Int J Med Robot Comput Assist Surg 16(4):e2105. https://doi.org/10.1002/rcs.2105
14. Kühnapfel U, Çakmak HK, Maaß H (2000) Endoscopic surgery training using virtual reality and deformable tissue simulation. Comput Graph 24(5):671–682. https://doi.org/10.1016/S0097-8493(00)00070-4
15. Qi D et al (2020) Surgeons with five or more actual cricothyrotomies perform significantly better on a virtual reality simulator. J Surg Res 252:247–254. https://doi.org/10.1016/j.jss.2020.03.021
16. Demirel D et al (2016) Virtual airway skills trainer (VAST) simulator. Stud Health Technol Inform 220:91–97
17. J. Farmer *et al.*, "Virtual rotator cuff arthroscopic skill trainer: Results and analysis of a preliminary subject study," In: *Proceedings of the 2020 the 4th International Conference on Information System and Data Mining*, New York, NY, USA, May 2020, pp. 139–143, https://doi.org/10.1145/3404663.3404673.
18. "Application of virtual reality technology in clinical medicine." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5622235/. Accessed 18 Nov 2020.
19. Luciano C, Banerjee P, DeFanti T (2009) Haptics-based virtual reality periodontal training simulator. Virtual Reality 13:69–85. https://doi.org/10.1007/s10055-009-0112-7
20. Samadbeik M, Yaaghobi D, Bastani P, Abhari S, Rezaee R, Garavand A (2018) The applications of virtual reality technology in medical groups teaching. J Adv Med Educ Prof 6(3):123–129
21. Sankaranarayanan G et al (2011) Validation of a novel laparoscopic adjustable gastric band simulator. Surg Endosc 25(4):1012–1018. https://doi.org/10.1007/s00464-010-1306-5
22. Lewis TM et al (2012) Can virtual reality simulation be used for advanced bariatric surgical training? Surgery 151(6):779–784. https://doi.org/10.1016/j.surg.2012.03.014
23. Stefanidis D, Heniford BT (2009) The formula for a successful laparoscopic skills curriculum. Arch Surg 144:77–82. https://doi.org/10.1001/archsurg.2008.528 (**discussion 82**)
24. Demirel D et al (2016) A hierarchical task analysis of cricothyroidotomy procedure for a virtual airway skills trainer (VAST) simulator. Am J Surg 212(3):475–484. https://doi.org/10.1016/j.amjsurg.2015.08.029
25. Hazey JW et al (2014) Why fundamentals of endoscopic surgery (FES)? Surg Endosc 28(3):701–703. https://doi.org/10.1007/s00464-013-3299-3
26. Hollnagel E (2003) Handbook of cognitive task design. CRC Press, Boca Raton
27. Cetinsaya B et al (2019) A task and performance analysis of endoscopic submucosal dissection (ESD) surgery. Surg Endosc 33(2):592–606. https://doi.org/10.1007/s00464-018-6379-6
28. Demirel D et al (2017) A hierarchical task analysis of shoulder arthroscopy for a virtual arthroscopic tear diagnosis and evaluation platform (VATDEP). Int J Med Robot Comput Assist Surg 13(3):e1799. https://doi.org/10.1002/rcs.1799
29. Corbett M, O'Connor P, Byrne D, Thornton M, Keogh I (2019) Identifying and reducing risks in functional endoscopic sinus surgery through a hierarchical task analysis. Laryngoscope Investig Otolaryngol 4(1):5–12. https://doi.org/10.1002/lio2.220
30. Myers EM, Anderson-Montoya BL, Fasano HT, Vilasagar S, Tarr ME (2019) Robotic sacrocolpopexy simulation model and associated hierarchical task analysis. Obstet Gynecol 133(5):905–909. https://doi.org/10.1097/AOG.0000000000003218
31. Wucherer P et al (2015) Vertebroplasty performance on simulator for 19 surgeons using hierarchical task analysis. IEEE Trans Med Imaging 34(8):1730–1737. https://doi.org/10.1109/TMI.2015.2389033

32. Lane R, Stanton NA, Harrison D (2006) Applying hierarchical task analysis to medication administration errors. Appl Ergon 37(5):669–679. https://doi.org/10.1016/j.apergo.2005.08.001
33. Phipps D, Meakin GH, Beatty PCW, Nsoedo C, Parker D (2008) Human factors in anaesthetic practice: insights from a task analysis. Br J Anaesth 100(3):333–343. https://doi.org/10.1093/bja/aem392
34. Al-Hakim L, Maiping T, Sevdalis N (2014) Applying hierarchical task analysis to improving the patient positioning for direct lateral interbody fusion in spinal surgery. Appl Ergon 45(4):955–966. https://doi.org/10.1016/j.apergo.2013.11.013
35. Kumar N, Thompson CC (2014) A novel method for endoscopic perforation management by using abdominal exploration and full-thickness sutured closure. Gastrointest Endosc 80(1):156–161. https://doi.org/10.1016/j.gie.2014.02.022
36. Joice P, Hanna GB, Cuschieri A (1998) Errors enacted during endoscopic surgery—a human reliability analysis. Appl Ergon 29(6):409–414. https://doi.org/10.1016/S0003-6870(98)00016-7
37. Gurudu SR, Ramirez FC (2013) Quality metrics in endoscopy. Gastroenterol Hepatol 9(4):228–233
38. Schulz CM, Krautheim V, Hackemann A, Kreuzer M, Kochs EF, Wagner KJ (2016) Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. BMC Anesthesiol 16(1):4. https://doi.org/10.1186/s12871-016-0172-7
39. Alqahtani A, Al-Darwish A, Mahmoud AE, Alqahtani YA, Elahmedi M (2019) Short-term outcomes of endoscopic sleeve gastroplasty in 1000 consecutive patients. Gastrointest Endosc 89(6):1132–1138. https://doi.org/10.1016/j.gie.2018.12.012
40. Asokkumar R, Babu MP, Bautista I, Lopez-Nava G (2020) The use of the overstitch for bariatric weight loss in Europe. Gastrointest Endosc Clin N Am 30(1):129–145. https://doi.org/10.1016/j.giec.2019.08.007
41. Zorron R et al (2017) Endoscopic sleeve gastroplasty using Apollo overstitch as a bridging procedure for superobese and high risk patients. Endoscopy. https://doi.org/10.1055/s-0043-119685
42. Barrichello S et al (2019) Endoscopic sleeve gastroplasty in the management of overweight and obesity: an international multicenter study. Gastrointest Endosc 90(5):770–780. https://doi.org/10.1016/j.gie.2019.06.013
43. de Moura DTH, de Moura EGH, Thompson CC (2019) Endoscopic sleeve gastroplasty: from whence we came and where we are going. World J Gastrointest Endosc 11(5):322–328. https://doi.org/10.4253/wjge.v11.i5.322
44. Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. Biophotonics Int 11(7):36–42