Validity evidence for the Simulated Colonoscopy Objective Performance Evaluation scoring system

Kristen D. Trinca, M.D.\textsuperscript{a}, Tiffany C. Cox, M.D.\textsuperscript{a}, Jonathan P. Pearl, M.D.\textsuperscript{a, b}, E. Matthew Ritter, M.D.\textsuperscript{a,*}

\textsuperscript{a}Norman M. Rich Department of Surgery, Uniformed Services University/Walter Reed National Medical Simulation Center, 4301 Jones Bridge Road, A-3010, Bethesda, MD 20814, USA; \textsuperscript{b}Department of Surgery, University of Maryland School of Medicine, Bethesda, MD, USA

\textbf{KEYWORDS:} Simulation; Colonoscopy; Objective performance; Flexible endoscopy; Endoscopy training

\textbf{Abstract}

\textbf{BACKGROUND:} Low-cost, objective systems to assess and train endoscopy skills are needed. The aim of this study was to evaluate the ability of Simulated Colonoscopy Objective Performance Evaluation to assess the skills required to perform endoscopy.

\textbf{METHODS:} Thirty-eight subjects were included in this study, all of whom performed 4 tasks. The scoring system measured performance by calculating precision and efficiency. Data analysis assessed the relationship between colonoscopy experience and performance on each task and the overall score.

\textbf{RESULTS:} Endoscopic trainees’ Simulated Colonoscopy Objective Performance Evaluation scores correlated significantly with total colonoscopy experience ($r = .61, P = .003$) and experience in the past 12 months ($r = .63, P = .002$). Significant differences were seen among practicing endoscopists, non-endoscopic surgeons, and trainees ($P < .0001$). When the 4 tasks were analyzed, each showed significant correlation with colonoscopy experience (scope manipulation, $r = .44, P = .044$; tool targeting, $r = .45, P = .04$; loop management, $r = .47, P = .032$; mucosal inspection, $r = .65, P = .001$) and significant differences in performance between the endoscopist groups, except for mucosal inspection (scope manipulation, $P < .0001$; tool targeting, $P = .002$; loop management, $P = .0008$; mucosal inspection, $P = .27$).

\textbf{CONCLUSIONS:} Simulated Colonoscopy Objective Performance Evaluation objectively assesses the technical skills required to perform endoscopy and shows promise as a platform for proficiency-based skills training.

Published by Elsevier Inc.

The objective assessment of technical skills to quantify competence is a growing trend in surgical education. Teaching flexible endoscopy skills is increasingly emphasized in surgical residency programs and among practicing surgeons. Low-cost, objective systems to assess and train flexible endoscopy skills are needed to ensure that surgical education programs are training endoscopists to a high level of proficiency. We have developed a non-computer-based physical colonoscopy simulator to meet those needs. Our model, Simulated Colonoscopy Objective Performance Evaluation (S.C.O.P.E), provides a realistic platform for simulating flexible endoscopy. The purpose of this study was to evaluate the ability of S.C.O.P.E. to assess the skills required for flexible endoscopy.

Assessing competence in colonoscopy has been widely researched, but variable standards for establishing competence currently exist. The Accreditation Council for...
Graduate Medical Education Residency Review Committee for Surgery recommends trainees perform 35 esophagogastroduodenoscopies and 50 diagnostic colonoscopies before the completion of residency, and the Society of American Gastrointestinal and Endoscopic Surgeons supports this as a minimum competence standard. In contrast, the American Society of Gastrointestinal Endoscopy has published a recommendation that trainees perform a minimum of 140 colonoscopies to be considered competent. Many have questioned these arbitrary numbers and tried to define competence with specific measurable criteria. Commonly evaluated skills include time to intubation of the cecum, frequency of cecal intubation within 30 minutes, withdrawal time, and adenoma detection rate. Using these objective criteria, several studies have shown that having performed 140 colonoscopies does not result in the successful completion of colonoscopy in ≥90% of cases. Recommendations from these studies range from 150 to 500 colonoscopies required to achieve competence. Such variability in recommendations demonstrates the inconsistency when procedural competence is judged by case numbers alone. A better solution for assessing competence in colonoscopy is clearly needed. The foundation of an endoscopy simulator is its ability to assess the level of expertise an endoscopist possesses. The EndoBubble task on the GI Mentor II colonoscopy simulator (Simbionix, Cleveland, OH) measures the psychomotor ability of an endoscopist. This task can differentiate novices (0 to 100 colonoscopies) from experts (≥1,000 colonoscopies) on the basis of speed and accuracy of performance. This simulated endoscopy task may also differentiate navigation skills of novices vis-à-vis intermediate-level endoscopists (100 to 250 colonoscopies). As such, it may be a useful tool for assessing endoscopic skills.

The benefit of virtual reality in simulating colonoscopy skills is the ability to generate innumerable clinical scenarios through computer programming and to track performance metrics automatically. With unit costs frequently exceeding $100,000, many training programs cannot afford a virtual reality platform, and some platforms require frequent high-level maintenance. Furthermore, there is limited tactile feedback in the programming of the virtual reality platform, and an appropriate amount of pressure and rotation of the scope are difficult to simulate.

The high cost, requirement for frequent maintenance, and haptic limitations of virtual reality simulators prompted us to develop our physical model for colonoscopy skills assessment. We aimed to develop a model that accurately assesses core endoscopy skills, is reproducible, is inexpensive to use and maintain, and poses no risk to patients. Our model, S.C.O.P.E., is a realistic physical model for the assessment of colonoscopy core skills. In this study, we sought to determine whether S.C.O.P.E. can reliably differentiate endoscopists of varying levels of skill and experience in 4 endoscopic tasks.

**Methods**

**Subjects**

Thirty-eight volunteer subjects (17 attending physicians and 21 residents or fellows) were recruited for this study and stratified on the basis of training. Practicing endoscopists (PEs; n = 11) consisted of gastroenterology attending physicians, minimally invasive gastrointestinal surgeons, and colorectal surgeons who all perform gastrointestinal endoscopy as a routine part of their practice. Nonendoscopic surgeons (NESs; n = 6) consisted of general surgeons who do not routinely perform endoscopy. The third and final group was endoscopic trainees (ETs; n = 21), consisting of general surgery residents (postgraduate years 1 to 5) and gastroenterology fellows in all 3 years of training. The ET group was further categorized into junior trainees (general surgery postgraduate years 1 to 4) and senior trainees (general surgery postgraduate year 5 and gastrointestinal fellows).

**Apparatus**

All trials were performed using an adult colonoscope (Olympus America, Center Valley, PA) with capabilities for insufflation, irrigation, and suction, connected to a standard Olympus imaging system and displayed on a 4:3 aspect ratio, standard-definition video monitor (Fig. 1). The colonoscopy simulator consists of a Kyoto Kagaku colonoscopy model (Kyoto Kagaku Co Ltd, Tokyo, Japan). Latex colon inserts were modified to create each of the 4 S.C.O.P.E. tasks and were secured in the colonoscopy model (Fig. 2). Each modified colon insert contains 2 S.C.O.P.E. tasks. One task is performed on insertion and the other on withdrawal. The S.C.O.P.E. tasks were created through a combination of task deconstruction, review of similar ongoing endoscopic skills assessment tools, and expert opinion. Key metrics were identified through analysis of pilot data on all reasonably

**Figure 1** Overview of S.C.O.P.E. setup with colonoscope and Kyoto Kagaku colonoscopy simulator.
quantifiable parameters that could be measured from the S.C.O.P.E. tasks. Only those parameters that showed significant differences between skill groups or were considered clinically important for patient safety were included in the scoring system. Additionally, scores for each of the tasks described below were normalized to the mean performance of an experienced endoscopist group to allow each task to contribute equally to the total S.C.O.P.E. score. Details of this process have been presented previously and are, at the time of this writing, under review for publication as a separate report.

Tasks

Scope manipulation (SM). This task requires the use of torque and tip deflection to move a shape in the colon to the 6 o’clock position and align that shape within an outline of the shape placed on the monitor screen. A shape is considered completed if it is held in alignment with the screen outline overlay for 5 seconds, as determined by the proctor. Alignment must be held for 5 seconds for each shape before continuing to the next shape. Time will be restarted if the edges of the colonic shape move outside of the lines of the overlay. The modified colon contains 10 shapes in various locations; the shapes are numbered and must be completed in the order they are numbered. Timing begins when the first shape is seen and ends when the last shape is completed or at a maximum of 10 minutes. This is scored by time and number of shapes accurately completed, with a penalty for omitting a shape (Fig. 3).

Tool targeting (TT). This task requires coordination to use biopsy forceps to contact a metal target. The task is to touch the biopsy forceps to the metal disc within the colonic lumen to complete the circuit. Successful contact is signified by the sound of a tone. To complete each disc, the forceps must remain in contact with the disc for 5 seconds. When contact between the forceps and the disc is disrupted, a different tone is heard. The disc must be visible on the screen and contact between the forceps and disc maintained for the full 5 seconds, or the timer will be restarted. Additionally, the forceps must be withdrawn out of the view after each completed disc. Timing begins when the first disc is seen and ends when the last disc is completed or at a maximum of 10 minutes. The modified colon contains 10 discs in various locations; the discs are numbered and must be completed in the order they are numbered. This is scored by time and number of discs successfully completed, with a penalty for omitting a disc (Fig. 4).
Loop management (LM). The purpose of this task is to perform scope navigation to correctly manage the formation of a standard alpha loop. The task is to insert the scope in the simulated rectum followed by navigation to the cecum through a redundant sigmoid (a total of 100 cm of modified colon). If a loop is formed, it must first be recognized and then successfully reduced to continue advancing the colonoscope. The task will begin in the left lateral decubitus position, but on request, the task can be repositioned to the supine, prone, or right lateral decubitus position. Additionally, abdominal pressure can be provided by the proctor where and when directed by the subject. Timing begins when the colonoscope is inserted into the rectum and ends when the “stop” sign is reached or at a maximum of 15 minutes. This is scored by time and number of polyps correctly identified, with penalties for failing to identify polyps, incorrectly indentifying other structures as polyps, and not performing retroflexion of the colonoscope in the rectum (Fig. 5).

Mucosal inspection (MI). This task is designed to indentify polyps on withdrawal of the colonoscope. Between 10 and 20 polyps are present throughout a length of simulated colon and rectum, requiring careful MI, including retroflexion. The colonoscope starts in the cecum, and MI is carried out during scope withdrawal all the way to the rectum, to include retroflexion. Timing begins when the colonoscope is withdrawn from the colon wall and ends when the colonoscope is removed from the anus or at a maximum of 15 minutes. This is scored by time and number of polyps correctly identified, with penalties for failing to identify polyps, incorrectly indentifying other structures as polyps, and not performing retroflexion of the colonoscope in the rectum.

Procedure

Each subject was read a standard script with instructions for performing each of the 4 S.C.O.P.E. tasks (SM, TT, LM, and MI). Each subject performed each of the 4 tasks. Insufflation, irrigation, and suction were available and could be used for assistance to complete each task. The previously identified key metrics for each task designed to reward precision and efficiency were recorded by a single investigator serving as test proctor. The same investigator served as proctor for each assessment. If the time to complete each task surpassed the predetermined cutoff time, a timing score of zero was given. Precision was measured by calculating a penalty score for each task. Then, these 2 scores were combined to obtain a total score for each task. Data were analyzed to assess the relationship between colonoscopy experience and performance on each of the tasks and the overall score. Performance differences among PEs, NESs, and ETs were also assessed for each task and the overall score.

Statistical analysis

Measures of central tendency were analyzed across groups using the Kruskal-Wallis test with Dunn’s posttest comparisons for nonparametric data or 1-way analysis of variance with Tukey-Kramer posttest comparison for parametric data. Proportions were compared using Fischer’s

### Table 1  Demographics

<table>
<thead>
<tr>
<th>Variable</th>
<th>ETs (n = 21)</th>
<th>P</th>
<th>NESs (n = 6)</th>
<th>P</th>
<th>PEs (n = 11)</th>
<th>P (ETs vs PEs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median years since completing residency (range)</td>
<td>0 (0–3)</td>
<td>&lt;.001</td>
<td>4.5 (2–34)</td>
<td>.21</td>
<td>9 (5–34)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Median total flexible gastrointestinal procedures performed (range)</td>
<td>50 (0–250)</td>
<td>.06</td>
<td>59 (50–2,500)</td>
<td>.03</td>
<td>1,100 (300–10,000)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Median procedures performed past 12 mo (range)</td>
<td>10 (0–94)</td>
<td>.03</td>
<td>0 (0–10)</td>
<td>.001</td>
<td>60 (5–300)</td>
<td>.004</td>
</tr>
<tr>
<td>Right-handedness</td>
<td>100%</td>
<td>NS</td>
<td>100%</td>
<td>NS</td>
<td>100%</td>
<td>NS</td>
</tr>
</tbody>
</table>

ET = endoscopic trainee; NES = nonendoscopic surgeon; PE = practicing endoscopist.
Exact test. Unpaired Welch’s $t$ tests and Mann-Whitney $U$ comparisons were also used as appropriate. Correlations were done with linear regression analysis. All analysis was done using GraphPad InStat version 3.10 32-bit for Windows (GraphPad Software, San Diego, CA).

**Results**

Results of comparisons of demographic data among the groups were as expected (Table 1). ETs were less clinically experienced than either NESs or PEs. The ET group had performed a similar number of total endoscopy procedures compared with the NES group, but both these groups were significantly less experienced than the PE group. Interestingly, the ET group had more recent endoscopic experience compared with the NES group, but again, both these groups’ recent experience fell significantly below that of the PE group. Although all subjects were right hand dominant, there was a significant difference in gender distribution between the NES and PE groups.

With respect to mean total S.C.O.P.E. score, significant differences were seen among PEs, NESs, and ETs (Fig. 6). Comparisons among groups on posttest analysis showed that the PE group significantly outperformed both the NES group and the junior ET group. The senior ET group also outperformed both the NES and junior ET groups. Performance in the PE and senior ET groups was similar.

When each of the 4 tasks was analyzed individually, statistically significant differences in performance among the endoscopist groups were demonstrated, with the exception of the MI task (SM, $P < .0001$; TT, $P = .002$; LM, $P = .0008$; MI, $P = .27$; Figs. 7–10). On the 3 statistically significant tasks, the PE group consistently and significantly outperformed the junior ET group (Figs. 7–9). The PE group also significantly outperformed the NES group in the SM task (Fig. 7). Similarly, the senior ET group significantly outperformed the junior ET group on the SM, TT, and LM tasks (Figs. 7–9) and also outperformed the NES group on SM. No significant differences were seen across any of the tasks between either the PE or senior ET group and the NES and junior ET groups (Figs. 7–10).
To assess whether the S.C.O.P.E. system accurately reflects the growing skill set of those on the flexible endoscopy learning curve, total colonoscopy experience and colonoscopy experience in the past 12 months were correlated with total S.C.O.P.E. score and the individual S.C.O.P.E. task scores among ETs. With respect to the total S.C.O.P.E. score, ETs’ performance correlated well and significantly with total colonoscopy experience ($r = .61, P = .003$). Similarly, colonoscopy experience in the past 12 months also showed a strong relationship ($r = .63, P = .002$). When the 4 tasks were analyzed individually, each showed a reasonably strong and statistically significant correlation with colonoscopy experience (SM, $r = .44, P = .044$; TT, $r = .45, P = .04$; LM, $r = .47, P = .032$; MI, $r = .65, P = .001$).

**Comments**

This study lends validity evidence to the construct that the S.C.O.P.E. platform objectively assesses the technical skills required to perform diagnostic flexible endoscopy. There was a statistically significant improvement in performance with increasing level of training. Our findings support the use of S.C.O.P.E. as a useful tool to assess endoscopic skills across all skill levels, and S.C.O.P.E. may ultimately be useful as training platform for high-stakes endoscopic skills tests and a tool to assess endoscopic competence for credentialing.

In this study, we assessed 4 tasks of endoscopic technique and used the simulator to objectively evaluate endoscopic performance at different levels of endoscopic training and experience. Each task and the total score demonstrated improvement with increased colonoscopy experience and were able to differentiate among groups expected to have different endoscopic skills. Colonoscopy experience in the past 12 months also contributed to improvement in performance, which supports the notion that endoscopic competence is not maintained without continued practice. We chose to run our linear regression analysis only on the ET group, because the S.C.O.P.E. platform is designed to represent the fundamental skills to perform diagnostic colonoscopy. Thus, it was not designed to differentiate subtle levels of advanced performance. Our
preliminary data would show that S.C.O.P.E. scores generally begin to plateau somewhere between 100 and 250 colonoscopies, which aligns well with many national guidelines and recommendations. The significant benefit lies in the ability of our platform to quantify this skill set, so that more gifted trainees are not required to perform more training than required, and less naturally gifted trainees can gain additional experience. This allows the most economic use of our increasingly limited training time.

No significant difference was seen in performance among the different levels of training on the MI task, including retroflexion in the rectum. In this task, the junior ET group had the most penalties for failure to retroflex, most likely because of inexperience with the colonoscope and not having the skill set to perform retroflexion. We believe that this task failed to differentiate among our groups for several reasons. First and foremost is the nature of the task itself. MI is known to be a task for which faster is not better. Unlike the virtual reality world, there is no low-cost way to quantify the percentage of the mucosa that is visualized. Thus, execution time for this task was not included in the scoring system. Our only proxy for thorough MI is the identification of polyps. Although some of our previous work has shown that novices miss significantly more polyps than experts (unpublished data), our simulated polyps may need to be more subtle and placed in more strategic locations to truly quantify expert levels of performance. This will likely be remedied as our model and its manufacturing process mature.

Assessing competence in flexible endoscopy is a contentious issue. Surgeons and gastroenterologists have long debated the merits of numbers-based assessment of competence. A tool such as S.C.O.P.E. may obviate the need for specific numbers of clinical cases and provide an objective measure of endoscopic skill. Using S.C.O.P.E., designation of proficiency and subsequent credentialing could be based on metrics supported by validity evidence.

Our study was limited by the fact that it was a single-institution study on a prototype simulation model. Furthermore, we do not know whether an expert score in S.C.O.P.E. translates into success on either clinical measures of improved performance, such as the Global Assessment of Gastrointestinal Endoscopic Skills or the Mayo Colonoscopy Skills Assessment Tool, or high-stakes tests, such as Fundamentals of Endoscopic Surgery. Additionally, S.C.O.P.E. is designed only to assess the technical skills required for diagnostic colonoscopy and does not assess the cognitive requirements, such as recognition of pathology and monitoring patient complications. Other assessment tools, such as the Mayo Colonoscopy Skills Assessment Tool and the cognitive portion of Fundamentals of Endoscopic Surgery, assess this domain, and we made a conscious decision to focus only on technical skills assessment as a first step in development. Finally, a single investigator served as the test proctor for all of the assessments included in this study, so the important aspect of interrater reliability cannot yet be evaluated. Additional studies incorporating multiple proctors and multiple centers are needed to determine the full utility of S.C.O.P.E. as a training and assessment tool.

Conclusions

Simulation continues to be an important tool for assessment and training of technical abilities of many surgical skills. Our study demonstrated that there are key metrics of endoscopy skills that separate experienced from inexperienced endoscopists. Our model using these metrics can provide a training platform for acquiring and assessing endoscopic skills.

References


