Analyzing Colonoscopy Training Learning Curves Using Comparative Hand Tracking Assessment

Keiran Barr[a], Lawrence Hookey[b], Tamas Ungi[a], Gabor Fichtinger[a], Matthew Holden[c]
[a]Laboratory for Percutaneous Surgery, School of Computing, Queen's University, Kingston, Canada; [b]Gastrointestinal Diseases Research Unit, Department of Medicine, Queen's University, Kingston, Canada; [c]School of Computer Science, Carleton University, Ottawa, Canada

ABSTRACT

A competency-based approach for colonoscopy training is particularly important, since the amount of practice required for proficiency varies widely between trainees. Though numerous objective proficiency assessment frameworks have been validated in the literature, these frameworks rely on expert observers. This process is time-consuming, and as a result, there has been increased interest in automated proficiency rating of colonoscopies. This work aims to investigate sixteen automatically computed performance metrics, and whether they can measure improvements in novices following a series of practice attempts. This involves calculating motion-tracking parameters for three groups: untrained novices, those same novices after undergoing training exercises, and experts. Both groups had electromagnetic tracking markers fixed to their hands and the scope tip. Each participant performed eight testing sequences designed by an experienced clinician. Novices were then trained on 30 phantoms and re-tested. The tracking data of these groups were analyzed using sixteen metrics computed by the Perk Tutor extension for Slicer. Statistical differences were calculated using a series of three t-tests, adjusting for multiple comparisons. All sixteen metrics were statistically different between pre-trained novices and experts, which provides evidence of their validity as measures of performance. Experts had fewer translational or rotational movements, a shorter and more efficient path, and performed the procedure faster. Pre- and post-trained novices did not significantly differ in average velocity, motion smoothness, or path inefficiency.

Keywords: Colonoscopy, electromagnetic tracking, Slicer, competence assessment, medical training, learning curves

1. INTRODUCTION

With colorectal cancer being the third most commonly diagnosed cancer worldwide, colonoscopies remain the standard of care for screening and intervention planning[1]. Learning to perform a colonoscopy requires extensive development of both cognitive and motor skills. Since the procedure is invasive and often uncomfortable, adequate practitioner training is critical for patient tolerance. In addition, competent practitioners reduce the chance of complications arising from the procedure[2]. Recently, medical education has been shifting away from time-based training in favor of competency-based training. A competency-based approach is particularly important for colonoscopy training, since the amount of training required for proficiency varies widely between trainees[3].

Previous research into improving colonoscopy training has explored the use of basic training tools to help novices attain fundamental skills before learning on more sophisticated phantoms or actual patients[4]. One of the more popular tools for these purposes is the wooden bench-top model. Initially proposed by Walsh et al.[5], it has since been validated as an effective practice tool for beginner trainees[4]. As with most competency-based assessments, proficiency is rated by experts using one of many scoring-based frameworks, the likes of which have been outlined by the Mayo Colonoscopy Skills Assessment Tool (MCSAT)[3], the Gastrointestinal Endoscopy Competency Assessment Tool (GiECAT)[6], the Assessment of Competency in Endoscopy (ACE)[7], and the Simulated Colonoscopy Objective Performance Evaluation (SCOPE)[8]. Though these frameworks aim to assess trainees objectively and are generally well-accepted, they rely on expert observers to perform the assessment. This can limit the utility of these systems for independent practice, and place further responsibilities on experienced educators and clinicians with limited time.

Research into the automation of clinical skill evaluation has gained considerable traction recently[9], particularly in the field of laparoscopic surgery. Aggarwal et al. used the kinematic data from the *da Vinci*™ laparoscopic system in the operating room to compare the path lengths, number of movements, and procedure duration of experienced and inexperienced surgeons[10]. Cristancho et al. compared pre-trained novices, post-trained novices, and experts on their laparoscopic skills during a mandarin orange dissection simulation task. The three groups were evaluated on the basis of their tool tip velocities, using a magnetic sensor attached to the tool handles[11].

However, compared to laparoscopy, the process of tracking a colonoscope tends to be relatively more complex, owing in part to the scope's flexible nature. This challenge has been addressed in various ways. Vilmann et al. used the Olympus ScopeGuide system, which comprises a colonoscope with electromagnetic (EM) coils built into the scope along its length, and a tracking device. They compared novices and experts on their path lengths[12]. Obstein et al. affixed 6 EM sensors along the exterior of the scope at regular intervals[13]. However, it is not just scope position data that is valuable in assessing proficiency; to address this, Konge et al. used a Microsoft Kinect™ to measure distance between the hands, in addition to the Olympus ScopeGuide system to track the scope[14]. In a study focusing specifically on repetitive strain injury of the right wrist as a result of performing the procedure, Mohankumar et al. used a magnetic tracking system with a glove on the operator's right hand[15]. Svendson et al. used an optical flow-based algorithm with a Microsoft Kinect™ to score colonoscopies[16], and Holden et al. used electromagnetic tracking of the wrists and elbows to compare novices and experts by their ranges of motion[17]. Nerup et al. used magnetic endoscope imaging to track the scope tip paths of novices and experts[18].

While these studies represent important steps toward automated trainee evaluation, they do not provide substantial insight into the process of learning nor how to track it over time. In this paper, we aim to address this problem by investigating sixteen performance metrics and using these metrics to measure improvements in novices following a training session. This will involve calculating motion-tracking parameters for three groups: untrained novices, those same novices after undergoing training exercises, and experts.

2. METHODS

We selected the wooden bench-top model to test fundamental motor skills necessary to perform a colonoscopy, since it is simple to use, easy to recreate, and widely accepted as an effective training method[4]. The model consists of a series of four consecutive three-by-three grids of holes. Plastic markers were inserted into the holes to show the participants which holes to enter. Sequences are named using 4 numbers to denote which holes should be entered in what order. For example, Figure 1 illustrates the sequence 4962.

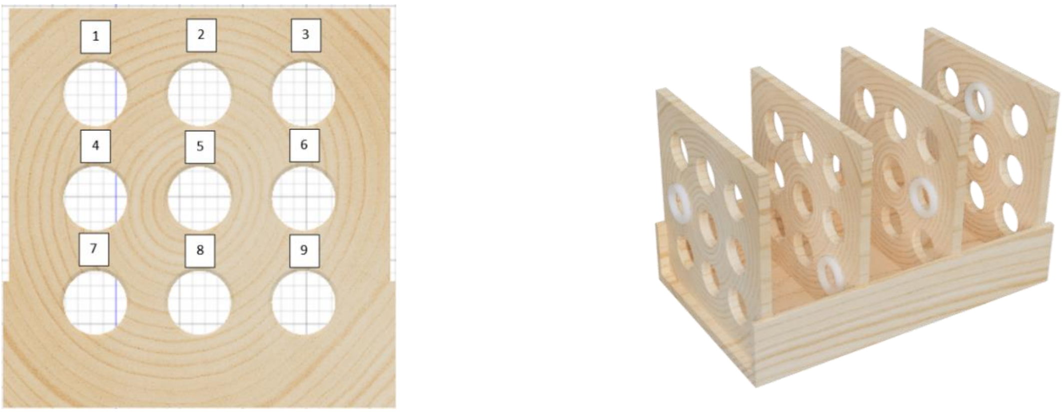

Figure 1. *Left*: front view of the bench-top model, with holes numbered; *Right*: uncovered phantom of sequence 4962

To monitor the position of the scope relative to the model, as well as the positions of the users' hands during the procedure, a spatial tracking system was used in conjunction with the model (Figure 2). We chose to use an electromagnetic (EM) tracking system (NDI 3D Guidance mid-range transmitter, Northern Digital Inc., Waterloo, Canada) for its reliability, as well as the non-metallic nature of the bench-top model. A tracking element was attached to the back of each of the participant's hands, and one was attached to the model to serve as a reference (3D Guidance model 800, Northern Digital

Inc., Waterloo, Canada). A fourth was inserted through one of the instrument channels of the colonoscope itself (3D Guidance model 180, Northern Digital Inc., Waterloo, Canada).

The EM tracking data were recorded into 3D Slicer (www.slicer.org), an open-source medical imaging program, using the PLUS software library (www.plustoolkit.org)[19]. Metrics were calculated on the tracking data using Perk Tutor (www.perktutor.org)[20], an open-source Slicer extension that enables for calculating performance metrics. For each sequence of the participant, we calculated metrics based on the scope tip, the right hand, and the left hand—compensating for participant handedness in our analysis. The metrics that we calculated were determined by expert consensus on relevant quantities for assessing skill in scope manipulation (Table 1).

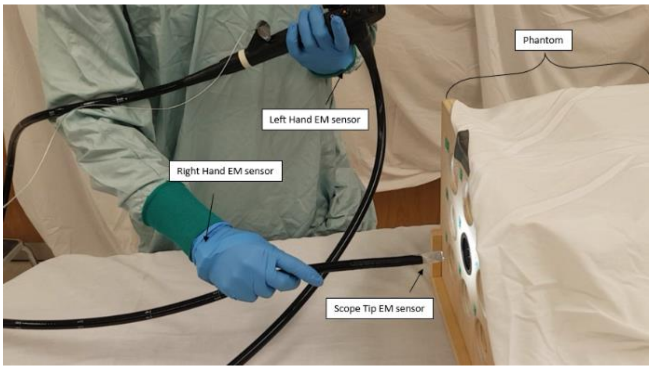

Figure 2. Experimental setup; shows three sensors, a covered phantom, and a participant

Table 1. List of all calculated metrics and their definitions

	Metric	Scope Tip	Hands	Definition
A	Path length (mm)	X	X	Total distance travelled by sensor
B	Average velocity (mm/s)	X	X	Average velocity of the sensor
C	Translational movements (#)	X	X	Number of times the translational velocity of the sensor exceeds 50 mm/s for more than 0.2 seconds
D	Rotational movements (#)	X	X	Number of times the rotational velocity of the sensor exceeds 50 degrees/s for more than 0.2 seconds
E	Motion smoothness (mm/s^3)	X		The total change in acceleration of the sensor
F	Depth perception (mm)	X		The total movement of the sensor in its axial direction
G	Path inefficiency (%)	X		Deviation of actual path length from optimal path length
H	Time elapsed (s)			Time taken between entering the first to final hole of the phantom

Two groups of participants were recruited—novice medical students (N=25) who had no experience with colonoscopies, and expert gastroenterologists (N=10) who had performed at least 1000 colonoscopies. Both experts and novices performed four unique testing sequences twice each, for a total of eight sequences. The order of these testing sequences was the same for all participants and was randomly determined. The novices then performed 30 training sequences that differed from the testing sequences. Finally, the novices were re-tested on the initial eight testing sequences, creating a third experimental group (N=24). This training paradigm is illustrated in Figure 3.

Group	Novices	Experts
Pre-training sequences	5877	5877
	2589	2589
	2588	2588
	6988	6988
	2588	2588
	5877	5877
	2589	2589
	6988	6988
	30 training sequences	
Post-training sequences	5877	
	2589	
	2588	
	6988	
	2588	
	5877	
	2589	
	6988	

Figure 3. Training paradigm, including the specific ordering of training and testing sequences. Note the three distinct experimental groups.

3. RESULTS AND DISCUSSION

Since the sequences varied in difficulty, metric values differ between sequences. Figure 4 shows the difference in number of scope tip translational movements between the three groups (pre-trained novices, post-trained novices, and experts) across the four testing sequences. Similarly, Figure 5 shows the difference in path length between the three groups. Similar patterns were observed for other metrics. Paired Student's t-tests were conducted to compare pre- and post-trained novices, and unpaired Welch's t-tests to compare pre-trained novices and experts, as well as post-trained novices and experts. This same framework was used for all 16 metrics, for a total of 48 comparisons. All tests were adjusted with Bonferroni correction (α=0.003).

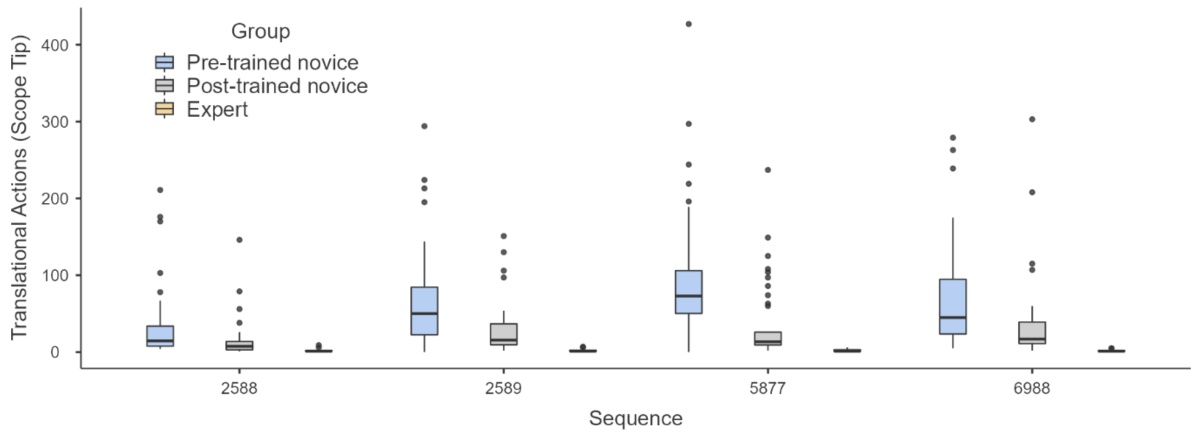

Figure 4. Translational actions of the scope tip for all three groups and all four sequences

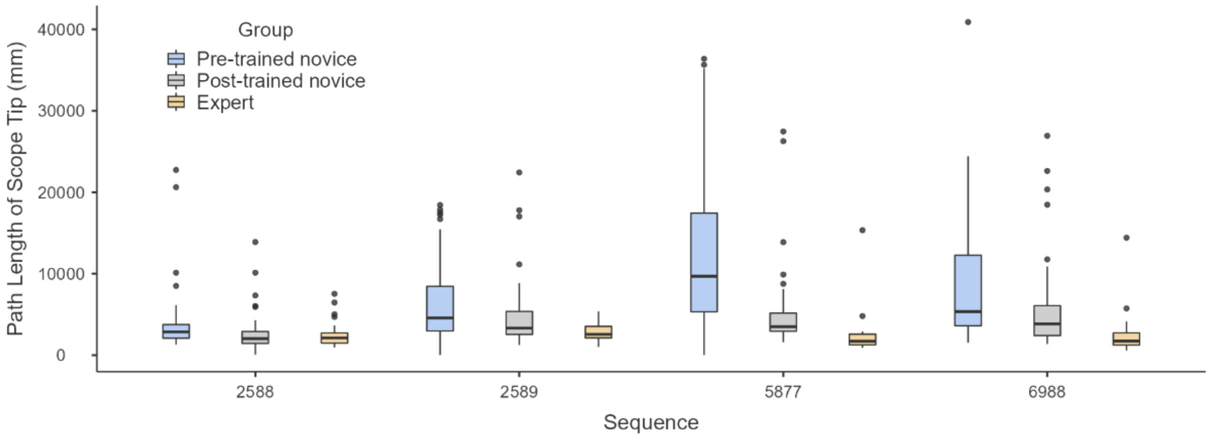

Figure 5. Path length of the scope tip for all three groups and all four sequences

Pre-trained novices and experts were statistically different across all metrics ($p<0.001$). Pre- and post-trained novices were statistically different for all metrics except average velocity of the dominant hand ($p=0.016$), path inefficiency of the scope tip ($p=0.009$), and motion smoothness of the scope tip ($p=0.305$). Post-trained novices and experts were statistically different for all metrics except average velocity of the dominant hand ($p=0.014$) and path inefficiency of the scope tip ($p=0.004$), though they did differ in their motion smoothness of the scope tip ($p<0.001$).

Experts exhibited less rotational and translational movements of both the scope and the hands. In addition, experts exhibited shorter and more efficient paths, as well as taking less time. Interestingly, experts had a higher average velocity than novices, though this may be because the novices had more motionless periods—likely due to their relative unfamiliarity with the equipment. Experts also had less smoothness of motion, though this metric has high uncertainty due to being a high order derivative of a noisy signal.

There are some limitations to be noted with this study. Firstly, while the wooden bench top model has been validated as an effective training tool, it is still far simpler than an actual colon. It is possible that the observed results might differ if the same study were to be conducted with colonoscopies on actual patients. In particular, the bench-top model does not account for important factors such as patient comfort or detection of abnormalities—it is purely a measure of gross colonoscope handling skills. As such, it does not paint a full picture of the overall skills of the individual performing the colonoscopy. Secondly, this study has low generalizability, as all the trials were performed at the same institution. Furthermore, our sample size is relatively small—particularly for experts (N=10). This study ideally should be continued to ensure repeatability. Finally, small sources of error might include equipment inaccuracies, subtle differences in the placement of the EM sensors, unnoticed movements of the field generator, or inaccurate calibrations.

Qualitatively, the wooden bench-top model was easy to work with, and a simple training paradigm was shown to be effective for improving the performance of novices. This is in line with the existing literature that validates its efficacy[4,21]. In an interesting study, Grover et al. found that novices progressed quicker when they began their colonoscopy training on the wooden bench-top model, before moving on to more difficult and sophisticated simulators[21]. Nerup et al. employed magnetic endoscopic imaging to visualize the scope tip paths of expert and novice participants. They found that the novices tended to get stuck on “problem areas” which the experts did not[18]. This might correlate with our findings of a longer path length and duration of procedure for novices. Furthermore, we provide evidence that our training paradigm can improve this common weakness of novices. Our findings that Perk Tutor can objectively measure colonoscopy performance metrics aligns with the results of Holden et al[17].

Importantly, this research has also documented that different sequences of the wooden bench-top model vary in difficulty. A systematic comparison of the difficulty of each sequence may be beneficial for future use of this model.

4. CONCLUSION

Evidence of validity of these sixteen metrics is shown by their ability to quantitatively differentiate between all three groups. These metrics were able to track improvements in performance over time: novices were found to perform better after training. Future studies are underway which aim to compare structured expert rating data with these computed metrics.

5. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

We investigated sixteen motion tracking metrics and evaluated their potential for monitoring improvements in novice colonoscopy practitioners.

6. ACKNOWLEGEMENTS

G. Fichtinger is supported as a Canada Research Chair is Computer-Integrated-Surgery.

REFERENCES

[1] Haggar, F. A., and Boushey, R. P. "Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors." *Clinics in colon and rectal surgery* 22.04 (2009): 191-197.
[2] Anderson, M. L., Pasha, T. M., and Leighton, J. A. "Endoscopic perforation of the colon: lessons from a 10-year study." *The American journal of gastroenterology* 95.12 (2000): 3418-3422.
[3] Sedlack, R. E. "The Mayo Colonoscopy Skills Assessment Tool: validation of a unique instrument to assess colonoscopy skills in trainees." *Gastrointestinal endoscopy* 72.6 (2010): 1125-1133.
[4] Walsh, C. M., et al. "Concurrent versus terminal feedback: it may be better to wait." *Academic Medicine* 84.10 (2009): S54-S57.
[5] Walsh, C. M., et al. "Bench-top versus virtual reality simulation training in Endoscopy: Expertise discrimination." *Canadian Journal Gastroenterology and Hepatology* 22.Suppl A (2008): 164.
[6] Walsh, C. M., et al. "Gastrointestinal endoscopy competency assessment tool: Reliability and validity evidence." *Gastrointestinal endoscopy* 81.6 (2015): 1417-1424.
[7] Sedlack, R. E., et al. "ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD." *Gastrointestinal endoscopy* 79.1 (2014): 1-7.
[8] Ritter, E. M., et al. "Simulated Colonoscopy Objective Performance Evaluation (SCOPE): A non-computer-based tool for assessment of endoscopic skills." *Surgical endoscopy* 27.11 (2013): 4073-4080.
[9] Levin, M., et al. "Automated methods of technical skill assessment in surgery: a systematic review." *Journal of surgical education* 76.6 (2019): 1629-1639.
[10] Aggarwal, R., et al. "An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room." Annals of surgery 245.6 (2007): 992.
[11] Cristancho, S. M., et al. "Feasibility of using intraoperatively-acquired quantitative kinematic measures to monitor development of laparoscopic skill." MMVR. 2007.
[12] Vilmann, A. S., et al. "Using computerized assessment in simulated colonoscopy: a validation study." Endoscopy International Open 8.06 (2020): E783-E791.
[13] Obstein, K. L., et al. "Evaluation of colonoscopy technical skill levels by use of an objective kinematic-based system." Gastrointestinal endoscopy 73.2 (2011): 315-321.
[14] Konge, L., et al. "Combining different methods improves assessment of competence in colonoscopy." Scandinavian Journal of Gastroenterology 52.5 (2017): 601-605.
[15] Mohankumar, D., et al. "Characterization of right wrist posture during simulated colonoscopy: an application of kinematic analysis to the study of endoscopic maneuvers." Gastrointestinal endoscopy 79.3 (2014): 480-489.
[16] Svendsen, M. B., et al. "Using motion capture to assess colonoscopy experience level." *World journal of gastrointestinal endoscopy* 6.5 (2014): 193.
[17] Holden, M. S., et al. "Objective assessment of colonoscope manipulation skills in colonoscopy training." *International journal of computer assisted radiology and surgery* 13 (2018): 105-114.
[18] Nerup, N., et al. "Assessment of colonoscopy by use of magnetic endoscopic imaging: design and validation of an automated tool." *Gastrointestinal endoscopy* 81.3 (2015): 548-554.

[19] Lasso, A., et al. "PLUS: open-source toolkit for ultrasound-guided intervention systems." *IEEE transactions on biomedical engineering* 61.10 (2014): 2527-2537.
[20] Ungi, T., et al. "Perk Tutor: an open-source training platform for ultrasound-guided needle insertions." *IEEE Transactions on Biomedical Engineering* 59.12 (2012): 3475-3481.
[21] Grover, S. C., et al. "Progressive learning in endoscopy simulation training improves clinical performance: a blinded randomized trial." *Gastrointestinal Endoscopy* 86.5 (2017): 881-889.