Objective Assessment of Colonoscope Manipulation Skills in Colonoscopy Training

Matthew S. Holden[1*], Chang Nancy Wang[2], Kyle MacNeil[1], Ben Church[1],
Lawrence Hookey[2], Gabor Fichtinger[1], Tamas Ungi[1]

[1]Laboratory for Percutaneous Surgery, School of Computing, Queen's University,
Kingston, ON, Canada

[2]Gastrointestinal Diseases Research Unit, Department of Medicine, Queen's
University, Kingston, ON, Canada

*Email: 72mh@queensu.ca

Abstract

Objective: Manipulation of the colonoscope is a technical challenge for novice clinicians which is best learned in a simulated environment. It involves the coordination of scope tip steering with scope insertion, using a rotated image as reference. The purpose of this work is to develop and validate a system which objectively assesses colonoscopy technical skills proficiency in an arbitrary training environment, allowing novices to assess their technical proficiency prior to real patient encounters.

Methods: We implemented a motion tracking setup to objectively analyze and assess the way operators perform colonoscopies, including an analysis of wrist and elbow joint motions. Subsequently, we conducted a validation study to verify whether our motion analysis could discriminate novice colonoscopists from experts. Participants navigated a wooden bench-top model using a standard colonoscope while their motions were tracked.

Results: The developed motion tracking setup allowed colonoscopists of varying levels of proficiency to have their colonoscope manipulation assessed, and was able to be operated by a trained non-technical operator. Novice operators had significantly greater median times (101.5s vs. 31.5s) and number of hand movements (62.0 vs. 21.5) than experts. Experts, however, spent a significantly greater proportion of time in extreme ranges of wrist and elbow joint motion than novices.

Conclusion: We have developed and implemented a hand and joint motion analysis system that is able to discriminate novices from experts based on objective measures of motion. These metrics could, thus, serve as proxies for technical proficiency during training.

Keywords: colonoscopy, simulation-based training, medical education, objective skill assessment

Introduction

Worldwide, medical education has been undergoing a shift from time-based to competency-based medical education (CBME). CBME allows trainees to progress through their curriculum only once they have achieved specific performance benchmarks. This contrasts the time-based model in which trainees were stuck in courses of fixed length, regardless of their progress or lack thereof. Most importantly, this prevents trainees who have not yet reached competency by the end of the course from moving into a clinical setting with real patients.

For colonoscopy, a competency-based curriculum is particularly important, as training correlates significantly with success (e.g. [1]) and reduces complications (e.g. [2]). For this training, it is essential to have a simulation environment in which trainees may practice, whether it be a physical, virtual, or mixed environment [3], with performance across different simulation environments tending to be consistent [4]. These environments are necessary because they allow trainees to learn the intervention's workflow and practice basic technical skills prior to real patient encounters.

The challenge of CBME is in continually tracking each trainee's individual learning curve as they progress through the curriculum, such that the point at which they achieve competency can be quantitatively determined. Many solutions for proficiency assessment are based on structured expert observation, such as global rating scales and procedure-specific checklists. The methods of Chak et al. [5], the Global Assessment of Gastrointestinal Endoscopic Skills (GAGES) [6], the Simulated Colonoscopy Objective Performance Evaluation (SCOPE) [7], the Assessment of Competency in Endoscopy (ACE) toolbox [8], and the Gastrointestinal Endoscopy Competency Assessment Tool (GiECAT) [9], [10] were each developed to facilitate assessment in colonoscopy. These approaches, however, require direct expert supervision, which is too time consuming for supervisors. Furthermore, there remain inconsistencies between experts, even using these structured methods [11].

Accordingly, there has been a recent shift towards methods for quantitative assessment of skill using time and position tracking measurements for CBME [12]. Overviews describing validity for many measures of objective proficiency assessment in colonoscopy on virtual simulators have been provided by Ansell et al. [13], Triantafyllou et al. [14], and Ekkelenkamp et al. [15]. For the Simbionix GI Mentor, construct validity for total time, efficiency, episodes of view loss, and episodes of patient pain has been shown, but metrics' discriminatory value depends on the difficulty of the case [16], [17]. MacDonald et al. [18] have shown validity for total time, percentage of time with patient pain, scope tip movement, and percentage of diseased region visualized on a predecessor of the CAE EndoVR. Likewise, Haycock et al. [19] demonstrated construct validity for total time, completion rate, use of variable stiffness, and sigmoid looping for the Olympus Endo TS-1. Plooy et al. [20] validated proficiency assessment measures including completion rates, total time and force applied to the colon on a physical model, the Kyoto Kagaku Colonoscope Training Model. Svendsen et al. [21] used hand motion capture on the same model and identified that the distance between hands differentiated novices from experts. Indeed, Telem et al. [22] have demonstrated that performance metrics in simulated colonoscopy translates well to successful colonoscopy in clinical practice.

The objective of this work is to develop a setup for the automatic, objective assessment of colonoscopy technical skills that is not specific to a particular model or simulation environment. To this end, we developed several objective measures of procedural efficiency, ergonomic efficiency, and joint motion efficiency. Subsequently, we validated their ability to discriminate novices from experts in a particular simulated colonoscopy environment, and integrated these metrics into an open-source software, where they could serve as proxies for technical proficiency during colonoscopy training.

A preliminary version of this work has been reported [23].

Methods

System Setup

We designed a motion tracking setup to record and assess hand, wrist, and elbow motions during the intervention. To this end, we attached electromagnetically tracked position and orientation sensors (3D Guidance trakSTAR, Northern Digital Inc., Waterloo, Canada) to operators' hands, forearms, and biceps and one sensor to the colon model (Fig. 1). The PLUS software library (www.plustoolkit.org) [24] was used to acquire and send the tracking data to the 3D Slicer software (www.slicer.org).

We developed an open-source colonoscopy performance analysis module based on the Perk Tutor platform (www.perktutor.org) [25] for image-guided interventions training. We implemented it in Python as a scripted module within the 3D Slicer environment. We elected to use Perk Tutor, as it allowed us to implement custom metrics that are specific to analyzing performance in colonoscopy. The module performs calibrations, records tracking data and computes the objective metrics of operator performance and efficiency. The module was designed to be usable by non-technical operators.

We elected to use free, open-source, and cross-platform software for this setup. The setup supports a variety of different position trackers and any colonoscope. Thus, this setup and results can be readily replicated at other medical training centres.

Objective Performance Metrics

To quantify colonoscope manipulation skills and performance during simulated colonoscopy, we used several motion efficiency metrics already integrated into the Perk Tutor extension: total time of procedure, total path length of hands [25], and the number of discrete hand motions [26]. Additionally, we implemented several objective metrics profiling wrist and elbow motions.

To compute wrist and elbow motion, calibration is required. For each wrist, we must determine the axes of flexion/extension motion and abduction/adduction (Fig. 2). For each elbow, we must determine the axes of flexion/extension and supination/pronation (Fig. 2). To determine these axes, the operator rotates their wrist or elbow in each rotational motion. The axes are separately computed over each rotational motion by calculating the eigenvector with the smallest associated eigenvalue of the matrix defined in Equation 1, where R_i is the instantaneous rotation matrix between the ith and $i+1$th transform from the hand sensor to the forearm sensor over a total of N recorded transforms.

$$\sum_{i=1}^{N-1}[(R_i - I)^T(R_i - I)] \qquad \text{Equation 1}$$

The axes of rotation can subsequently be used to determine coordinate systems on the hand and on the forearm that are aligned with the wrist or elbow's axes. The angles of wrist and elbow rotation about each axis can thus be determined by the rotation between the aligned coordinate systems on the hand and forearm and the forearm and bicep, respectively (Fig. 3).

For each joint, we measured the number of times the joint entered extreme ranges of motion, as well as the total time spent in extreme ranges of motion. For wrists, this was with respect to flexion/extension and adduction/abduction. For elbows, this was with respect to flexion/extension and supination/pronation. A joint was considered to be in an extreme range of motion when the joint's angle exceeded 20% of the total range of motion for the operator in either direction, following the protocol outlined by Mohankumar et al. [27]. The total ranges of motion were computed from the prior calibration. These metrics are proxies for the ergonomic stress and strain exerted on each joint, and assess the ergonomic efficiency of the operator.

We implemented additional metrics to assess the efficiency of joint motions. One metric computes the number of discrete rotational motions for each joint (the rotational analog of discrete hand motions). Each discrete motion is defined as a period of rotation of at least 50°/s about any axis, delineated by a period of at least 0.2s of rest. The other metric measures the cumulative angle of rotation through which each joint was rotated (the rotational analog of total path length). This quantity is computed by summing all the instantaneous angles of rotation.

Finally, we implemented a metric to assess the operator's arm joint coordination. This metric calculates the correlation between the wrist and elbow rotation relative to the neutral position (Equation 2), where θ is a function that computes the angle of rotation from a rotation matrix, W_i is the ith rotation matrix from hand to forearm, W' is the rotation matrix from hand to forearm in the neutral position, E_i is the ith rotation matrix from forearm to bicep, and E' is the rotation matrix from forearm to bicep in the neutral position, all over a sequence of N frames.

$$\frac{\sum_{i=1}^{N}\left[\theta\left({W_i}^{-1}W'\right)\theta\left({E_i}^{-1}E'\right)\right]}{\sqrt{\sum_{i=1}^{N}\left[\theta\left({W_i}^{-1}W'\right)^2\right]\sum_{i=1}^{N}\left[\theta\left({E_i}^{-1}E'\right)^2\right]}}$$

Equation 2

While success rate is an important consideration, these time and efficiency metrics provide valuable information. In particular, they indicate colonoscope manipulation skill and understanding. This is important because greater colonoscope manipulation skill translates to reduced operating times and patient discomfort. Furthermore, when the trainee reaches the "unconscious competence" stage of learning colonoscope manipulation, it allows them to concentrate on other aspects of the procedure such as image interpretation, diagnosis, and patient management.

Validation Study

Both the standard efficiency metrics and the joint motion metrics were tested for evidence of construct validity (whether they discriminate experts from novices) by conducting a study on simulated colonoscopies (Fig. 4). Twenty-two novice and eight expert colonoscopists were recruited to perform simulated colonoscopies. All participants were right-handed. The novice group consisted of medical students with no prior simulated or clinical colonoscopy experience. The expert group consisted of staff gastroenterologists who perform at least 200 colonoscopies per year and have at least five years of experience. Participants performed the simulated colonoscopies on a previously validated wooden bench-top model (Fig. 5). The model is intended for teaching essential colonoscope manipulation techniques to medical trainees with no previous colonoscopy experience, before they move on to more difficult and realistic models. This model has been extensively validated, demonstrating its efficacy as a training tool for low-level trainees [28]. Furthermore, its flexible design allows participants to navigate different training sequences to practice different maneuvers involved in colonoscopy.

First, the electromagnetic sensors were attached securely to the participant's hands, forearms, and elbows. The participant was subsequently asked to perform the calibration exercises (two motions for each wrist and elbow, totaling eight exercises), and the software module automatically computed the calibrations. Next, the participant was assigned five practice navigation sequences to familiarize themselves with the colonoscope and the wooden bench-top model. Finally, each participant was assigned the same set of eight navigation sequences: four unique sequences each performed twice in random order. Participants were given a maximum of eight minutes to complete each sequence and were stopped if they had not completed the sequence within this time. The order of sequences was the same for all participants. The eight navigation sequences were tracked and analyzed using the colonoscopy software module.

Statistical Analysis

The difference in the proportion of completed trials between the novice and expert group was tested using Fisher's Exact test (α=0.05). Performance metric data were tested for normality using the Jarque-Bera test and found to be non-normally distributed. Differences between the novice and expert group were tested using the Mann-Whitney U test, with the Bonferroni correction for multiple tests (α=0.0017). The reported effect sizes are computed non-parametrically using Cliff's Δ, using the interpretation scheme outlined by Romano et al. [29].

Results

The custom colonoscopy software module was successfully able to record, calibrate, and analyze the tracking data from the simulated colonoscopy procedures. The module is available as an open-source module in the 3D Slicer environment, available for use without restriction on Linux, Mac OSX, and Windows. Using the module, the complete calibration can be performed in less than five minutes and the performance metrics are computed automatically. The software was successfully used by several different operators with varying technical backgrounds.

Experts successfully completed the trials within the allotted eight minutes significantly more often than novices (100% vs. 93%, p=0.02). Overall, experts exhibited better scores for all standard performance efficiency metrics, with significantly lower median time, hand path length, and number of discrete hand motions than novices (Fig. 6, Table 1), with large effect size for each performance efficiency metric. This demonstrates validity for the proposed performance efficiency metrics as discriminators of proficiency.

For each of right wrist flexion/extension, right wrist abduction/adduction, right elbow flexion/extension, and right elbow supination/pronation, novices entered extreme ranges of motion significantly more often than experts during a given simulated procedure (Table 1), medium to large effect size. This pattern, however, did not hold for the left side, where only left wrist abduction/adduction was significantly greater for novices. Furthermore, novices spent a significantly greater amount of time in extreme ranges of motion than experts for the right side, and for the left side except for left elbow flexion/extension. These results indicate that experts exerted less ergonomic stress and strain on their right wrist and elbow joints.

When time was factored out, however, experts appeared significantly less ergonomically efficient for the left side. Experts entered extreme ranges of motion significantly more frequently than novices for left wrist flexion/extension, left wrist adduction/abduction, left elbow flexion/extension, and left elbow supination/pronation. Additionally, experts spent a greater proportion of time in extreme ranges of motion for left wrist flexion/extension and left elbow flexion/extension. For the right side, experts entered extreme ranges of motion significantly more frequently for only right wrist flexion/extension and effect size was smaller (Δ=0.267).

Novices had more total rotation and rotational motions in both the right wrist and elbow than experts, but there was no difference in the left wrist or elbow (Table 1). Novices exhibited higher joint coordination in the left arm, but joint coordination was not significantly different in the right arm.

Discussion

Overall, experts completed the procedure more often, more efficiently, and exerted less ergonomic stress and strain on their joints. When time was factored out, however, experts entered extreme ranges of motion significantly more frequently and spent a greater proportion of time in extreme ranges of motion for the left side. We conjecture this is because novices spend more time contemplating how to manipulate the colonoscope correctly, accumulating time in non-extreme ranges of motion, and experts are more comfortable adjusting their left side to facilitate insertion during difficult procedures.

This study is not without limitations. Participants performed simulated colonoscopies on a wooden bench-top model. During these simulated colonoscopies, participants needed not interpret images, identify anatomy, nor manage the patient. It is unclear how performance on the model transfers to proficiency in an operating environment; however, there is evidence that the technical skills are transferable [22]. Furthermore, to ensure data integrity for validation, we did not test if the participants could use the system without a trained operator. There could be an “audience effect” causing participants to perform differently than they would in a self-guided setting [30]. Further study is required to understand how these results translate to a completely self-guided medical training scenario. Additionally, under the ranges of motion convention of Mohankumar et al. [27], if the joint angle exceeded 20% of the full range of joint motion, it was considered extreme. This threshold, however, was not determined anatomically. For our study, this definition of extreme range of motion is perhaps too strict, as each joint is in an extreme range of motion the majority of the time (Fig. 8). A less strict or fuzzy definition of this threshold may better discern experts from novices.

A challenge with validating performance metrics is determining a gold-standard against which to validate them. In this study, we used experience as a reference for validation. This cannot, however, account for anomalous performances (i.e. experts performing poorly or novices performing well by happenstance) and “bad habits” developed by experts. Although some experts undoubtedly exhibit certain undesirable habits, our expert group appears big enough to have diluted these. In addition, all experts had completed the hands-on Canadian Association of Gastroenterology Skills Enhancement in Endoscopy©, which addresses such habits and corrects behaviour. As a result, the majority of experts would be classified as “torque steerers” as opposed to “wheel steerers”, as conversion to this technique is one of the main tenets of the course.

The proposed system shows promise for use as one component in an experimental colonoscopy training curriculum. This would allow trainees’ improvement relative to performance benchmarks to be monitored as they progress through the curriculum. Providing such feedback in an objective manner is an important component of the newly evolving competency-based medical education paradigm and providing it automatically allows trainees to practice their technical skills without expert supervision. Furthermore, the analysis of operators’ ergonomic efficiency could be incorporated into a training program which teaches the ergonomic use of the colonoscope, which would contribute to the prevention of workplace injuries. Although colonoscopy is one of the highest-risk interventions for injury due to repetitive motion strain, the proposed measures of ergonomic efficiency may be useful in other applications, such as ultrasound-guided needle interventions or laparoscopy. Because the ranges of motion analysis is not colonoscopy-specific and the setup is readily replicable, it could be reused for the analysis of ergonomic efficiency in other minimally invasive interventions.

The results achieved in this study are consistent with other works on performance assessment in colonoscopy on physical simulators, which report that experts have greater rates of success and efficiency compared to novices [20], [21]. This is likewise consistent with results on purely virtual simulators [13], [14], [15]. Our study has shown that experts also exert less ergonomic stress and strain on their wrists and elbows. Given the relation between performance on colonoscopy simulators and patients, developing such measures for monitoring performance is critically important [22].

In future work, we suggest that the tip of the colonoscope could be tracked in addition to the hands, forearms, and biceps [31]. This would allow the system to analyze the method in which each operator navigates the colonoscope, which can provide further information on operator performance [32]. Using such measures of efficiency and accuracy of the colonoscope's path in combination with measures of hand and arm efficiency could lead to even more reliable assessment [33]. These performance metrics and others could be readily added to the system as custom performance metrics within Perk Tutor platform. We envision using these metrics to monitor trainees' performance as they practice on increasingly higher fidelity models as a step towards overall competence in colonoscopy. Furthermore, tracking information could be used to provide the operator with various forms of guidance, which may improve the training process for novice operators.

Conclusion

We have implemented a hardware and software setup intended to assess technical skills in colonoscopy, which can be readily replicated and adapted to any simulation environment. This is intended to address the paucity in objective assessment methods for simulation-based colonoscopy training. The proposed measures of overall efficiency, ergonomic efficiency, and joint efficiency successfully discriminated experts from novices in a simulated colonoscopy study, showing evidence of validity for indicating operator proficiency. In the future, we envision that a system capable of automatically computing such metrics as proxies for technical proficiency could be integrated into a competency-based colonoscopy training curriculum for monitoring trainee progress.

Acknowledgements

Matthew S. Holden is supported by the Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada. Gabor Fichtinger is supported as a Cancer Care Ontario Research Chair in Cancer Imaging. Financial support was received from the SEAMO Educational Innovation and Research Fund. This work was financially supported as a Collaborative Health Research Project (CHRP #127797), a joint initiative between the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institutes of Health Research (CIHR).

Compliance with Ethical Standards

Conflict of Interest

All authors declare that they have no conflict of interest.

Ethical Approval

All procedures in this study involving human participants were performed in accordance with the ethical standards of the institution, and were approved by the research ethics board at Queen's University. This study does not contain any procedures involving animals.

Informed Consent

All participation was voluntary, and written informed consent was obtained from all participants.

References

[1] J. B. Marshall, "Technical proficiency of trainees performing colonoscopy: A learning curve," *Gastrointestinal Endoscopy ,* vol. 42, pp. 287-291, 1995.

[2] M. L. Anderson, T. M. Pasha and J. A. Leighton, "Endoscopic perforation of the colon: lessons from a 10-year study," *Am. J. Gastroenterol.,* vol. 95, pp. 3418-3422, Dec 2000.

[3] A. Hill, M. S. Horswill, A. M. Plooy, M. O. Watson, R. Karamatic, T. A. Basit, G. M. Wallis, S. Riek, R. Burgess-Limerick and D. G. Hewett, "Assessing the realism of colonoscopy simulation: the development of an instrument and systematic comparison of 4 simulators," *Gastrointest. Endosc.,* vol. 75, pp. 631-640, Mar 2012.

[4] L. Preisler, M. B. Svendsen, N. Nerup, L. B. Svendsen and L. Konge, "Simulation-based training for colonoscopy: establishing criteria for competency," *Medicine (Baltimore),* vol. 94, p. e440, Jan 2015.

[5] A. Chak, G. S. Cooper, E. W. Blades, M. Canto and M. V. Sivak, "Prospective assessment of colonoscopic intubation skills in trainees," *Gastrointest. Endosc.,* vol. 44, pp. 54-57, Jul 1996.

[6] M. C. Vassiliou, P. A. Kaneva, B. K. Poulose, B. J. Dunkin, J. M. Marks, R. Sadik, G. Sroka, M. Anvari, K. Thaler, G. L. Adrales, J. W. Hazey, J. R. Lightdale, V. Velanovich, L. L. Swanstrom, J. D. Mellinger and G. M. Fried, "Global Assessment of Gastrointestinal Endoscopic Skills (GAGES): a valid measurement tool for technical skills in flexible endoscopy," *Surg Endosc,* vol. 24, pp. 1834-1841, Aug 2010.

[7] E. M. Ritter, T. C. Cox, K. D. Trinca and J. P. Pearl, "Simulated Colonoscopy Objective Performance Evaluation (SCOPE): a non-computer-based tool for assessment of endoscopic skills," *Surg Endosc,* vol. 27, pp. 4073-4080, Nov 2013.

[8] R. E. Sedlack, W. J. Coyle, K. L. Obstein, M. A. Al-Haddad, G. Bakis, J. A. Christie, R. E. Davila, B. DeGregorio, C. J. DiMaio, B. K. Enestvedt, J. Jorgensen, D. K. Mullady and L. Rajan, "ASGEś assessment of competency in endoscopy evaluation tools for colonoscopy and EGD," *Gastrointest. Endosc.,* vol. 79, pp. 1-7, Jan 2014.

[9] C. M. Walsh, S. C. Ling, N. Khanna, M. A. Cooper, S. C. Grover, G. May, T. D. Walters, L. Rabeneck, R. Reznick and H. Carnahan, "Gastrointestinal Endoscopy Competency Assessment Tool: development of a procedure-specific assessment tool for colonoscopy," *Gastrointest. Endosc.,* vol. 79, pp. 798-807, May 2014.

[10] C. M. Walsh, S. C. Ling, N. Khanna, S. C. Grover, J. J. Yu, M. A. Cooper, E. Yong, G. C. Nguyen, G. May, T. D. Walters, R. Reznick, L. Rabeneck and H. Carnahan, "Gastrointestinal Endoscopy Competency Assessment Tool: reliability and validity evidence," *Gastrointest. Endosc.,* vol. 81, pp. 1417-1424, 2015.

[11] R. Aggarwal, T. P. Grantcharov and A. Darzi, "Framework for systematic training and assessment of technical skills," *J. Am. Coll. Surg.,* vol. 204, pp. 697-705, Apr 2007.

[12] C. E. Reiley, H. C. Lin, D. D. Yuh and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical Endoscopy,* vol. 25, pp. 356-366, 2011.

[13] J. Ansell, J. Mason, N. Warren, P. Donnelly, N. Hawkes, S. Dolwani and J. Torkington, "Systematic review of validity testing in colonoscopy simulation," *Surg Endosc,* vol. 26, pp. 3040-3052, Nov 2012.

[14] K. Triantafyllou, L. D. Lazaridis and G. D. Dimitriadis, "Virtual reality simulators for gastrointestinal endoscopy training," *World J Gastrointest Endosc,* vol. 6, pp. 6-12, Jan 2014.

[15] V. E. Ekkelenkamp, A. D. Koch, A. de Man and R. and E. J. Kuipers, "Training and competence assessment in GI endoscopy: a systematic review," *Gut,* vol. 65, pp. 607-615, Apr 2016.

[16] A. D. Koch, S. N. Buzink, J. Heemskerk, S. M. Botden, R. Veenendaal, J. J. Jakimowicz and E. J. Schoon, "Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy," *Surg Endosc,* vol. 22, pp. 158-162, Jan 2008.

[17] R. Fayez, L. S. Feldman, P. Kaneva and G. M. Fried, "Testing the construct validity of the Simbionix GI Mentor II virtual reality colonoscopy simulator metrics: module matters," *Surgical Endoscopy,* vol. 24, pp. 1060-1065, 2010.

[18] J. MacDonald, J. Ketchum, R. G. Williams and L. Q. Rogers, "A lay person versus a trained endoscopist: can the preop endoscopy simulator detect a difference?," *Surg Endosc,* vol. 17, pp. 896-898, Jun 2003.

[19] A. V. Haycock, P. Bassett, J. Bladen and S. Thomas-Gibson, "Validation of the second-generation Olympus colonoscopy simulator for skills assessment," *Endoscopy,* vol. 41, pp. 952-958, Nov 2009.

[20] A. M. Plooy, A. Hill, M. S. Horswill, A. S. Cresp, M. O. Watson, S. Y. Ooi, S. Riek, G. M. Wallis, R. Burgess-Limerick and D. G. Hewett, "Construct validation of a physical model colonoscopy simulator," *Gastrointest. Endosc.,* vol. 76, pp. 144-150, Jul 2012.

[21] M. B. Svendsen, L. Preisler, J. G. Hillingsoe, L. B. Svendsen and L. Konge, "Using motion capture to assess colonoscopy experience level," *World J Gastrointest Endosc,* vol. 6, pp. 193-199, May 2014.

[22] D. A. Telem, D. W. Rattner and D. W. Gee, "Endoscopic simulator curriculum improves colonoscopy performance in novice surgical interns as demonstrated in a swine model," *Surgical Endoscopy,* vol. 28, pp. 1494-1499, 2014.

[23] K. MacNeil, C. N. Wang, M. S. Holden, T. Ungi, G. Fichtinger and L. Hookey, "System for objectively evaluating colonoscopy procedural skills using motion analysis," in *14th Annual Imaging Network Ontario Symposium*, 2016.

[24] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi and G. Fichtinger, "PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems," *IEEE Transactions on Biomedical Engineering,* vol. 61, pp. 2527-2537, Oct 2014.

[25] T. Ungi, D. Sargent, E. Moult, A. Lasso, C. Pinter, R. C. McGraw and G. Fichtinger, "Perk Tutor: an open-source training platform for ultrasound-guided needle insertions," *IEEE Trans Biomed Eng,* vol. 59, pp. 3475-81, Dec 2012.

[26] R. McGraw, T. Chaplin, C. McKaigney, L. Rang, M. Jaeger, D. Redfearn, C. Davison, T. Ungi, M. Holden, C. Yeo, Z. Keri and G. Fichtinger, "Development and Evaluation of a Simulation-

based Curriculum for Ultrasound-guided Central Venous Catheterization," *CJEM,* pp. 1-9, May 2016.

[27] D. Mohankumar, H. Garner, K. Ruff, F. C. Ramirez, D. Fleischer, Q. Wu and M. Santello, "Characterization of right wrist posture during simulated colonoscopy: an application of kinematic analysis to the study of endoscopic maneuvers," *Gastrointestinal Endoscopy ,* vol. 79, pp. 480-489, 2014.

[28] C. M. Walsh, S. C. Ling, C. S. Wang and H. Carnahan, "Concurrent versus terminal feedback: it may be better to wait," *Acad Med,* vol. 84, pp. S54--57, Oct 2009.

[29] J. Romano, J. D. Kromrey, J. Coraggio and J. Skowronek, "Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen’s d for evaluating group differences on the NSSE and other surveys," in *annual meeting of the Florida Association of Institutional Research*, 2006.

[30] C. F. Bond and L. J. Titus, "Social facilitation: a meta-analysis of 241 studies," *Psychol Bull,* vol. 94, pp. 265-292, Sep 1983.

[31] N. Nerup, L. Preisler, M. B. S. Svendsen, L. B. Svendsen and L. Konge, "Assessment of colonoscopy by use of magnetic endoscopic imaging: design and validation of an automated tool," *Gastrointestinal Endoscopy,* vol. 81, pp. 548-554, 2015.

[32] L. Preisler, S. a. M. B. Svendsen, B. S?ndergaard, L. Brink, T. Nordentoft, L. B. Svendsen and L. Konge, "Automatic and unbiased assessment of competence in colonoscopy: exploring validity of the Colonoscopy Progression Score CoPS," *Endosc Int Open,* vol. 4, pp. E1238-E1243, Dec 2016.

[33] L. Konge, M. B. S. Svendsen, L. Preisler, L. B. Svendsen and Y. S. Park, "Combining different methods improves assessment of competence in colonoscopy," *Scandinavian Journal of Gastroenterology,* vol. 52, pp. 601-605, 2017.

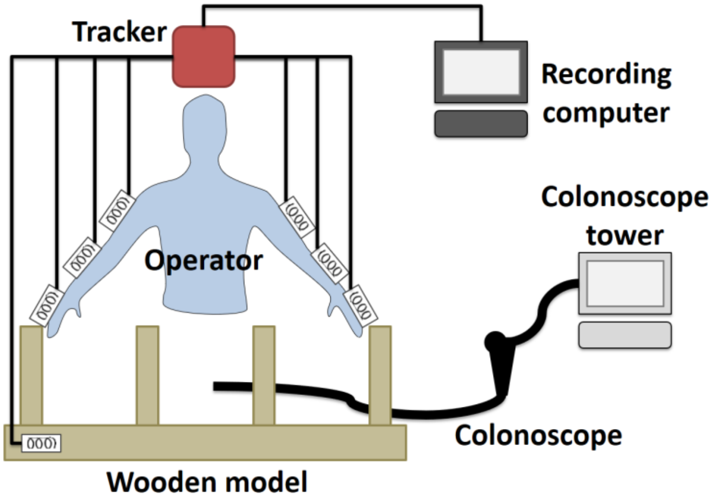

Fig. 1. Diagram of the hardware setup. Sensors are placed on the operator’s hands, forearm, and biceps. One sensor is also placed on the wooden bench-top model.

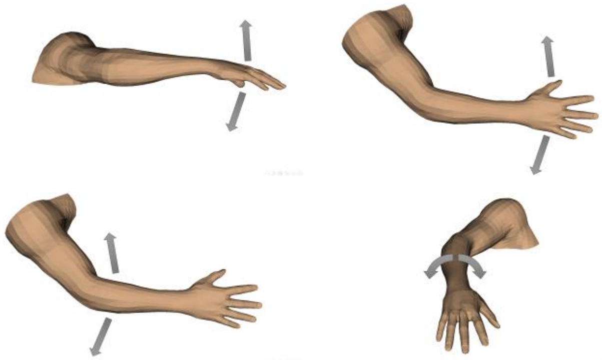

Fig. 2. Wrist flexion/extension (top-left), wrist abduction/adduction (top-right), elbow flexion/extension (bottom-left), and elbow supination/pronation (bottom-right) for the right side. Motions for the left side are analogous.

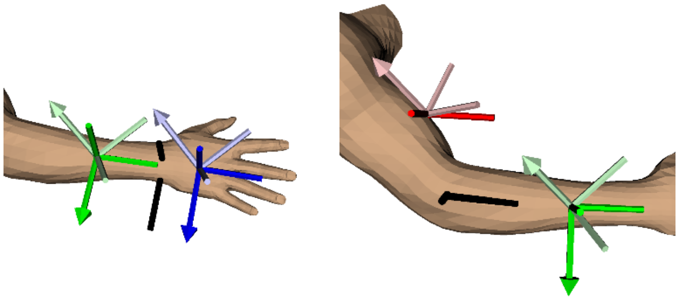

Fig. 3. Calibration for the wrist (left) and the elbow (right). The pale-colored axes represent the uncalibrated coordinate frames given by the tracking system and the bright-colored axes represent the calibrated coordinate frames. The black vectors represent the axes of rotation for wrist flexion/extension and abduction/adduction and elbow flexion/extension and supination/pronation.

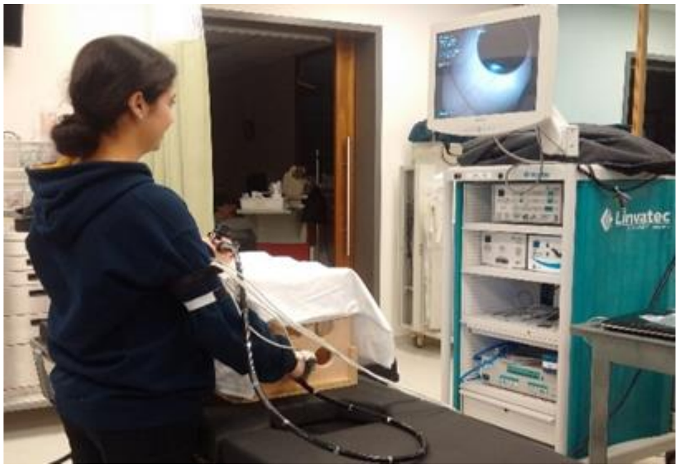

Fig. 4. A novice participant performing a simulated colonoscopy on the wooden bench-top model. Sensors are strapped to the hand, forearm, and bicep.

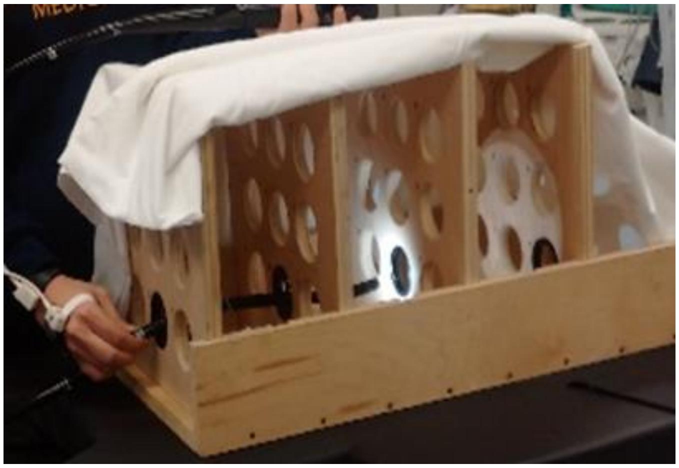

Fig. 5. Close-up view of the wooden bench-top model. The black rings indicate the holes through which the operator must navigate the colonoscope.

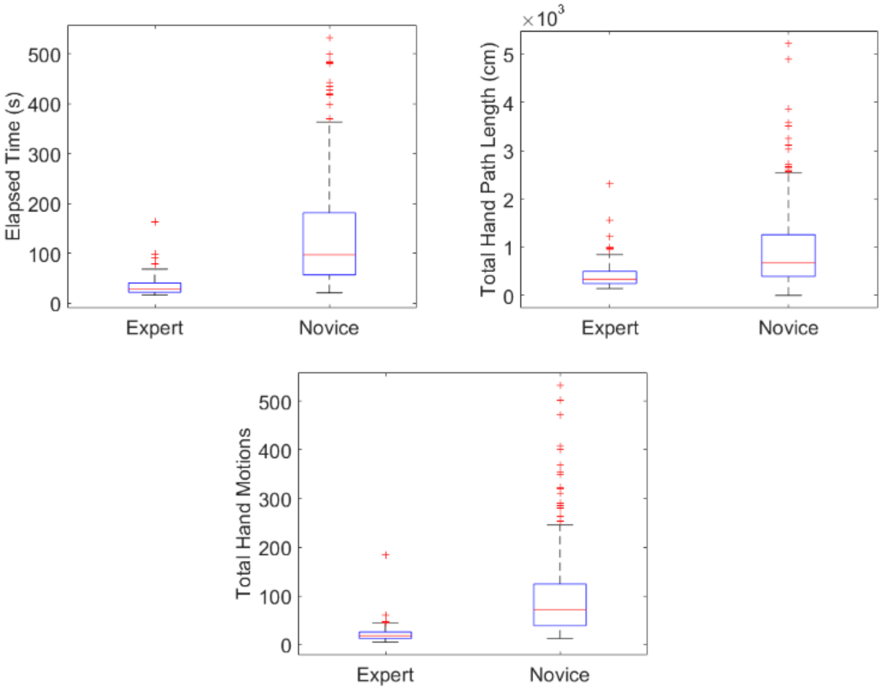

Fig. 6. Expert and novice total time (top-left), hand path length (top-right), and number of hand motions (bottom) for the recorded simulated colonoscopy sequences performed on the wooden bench-top model.

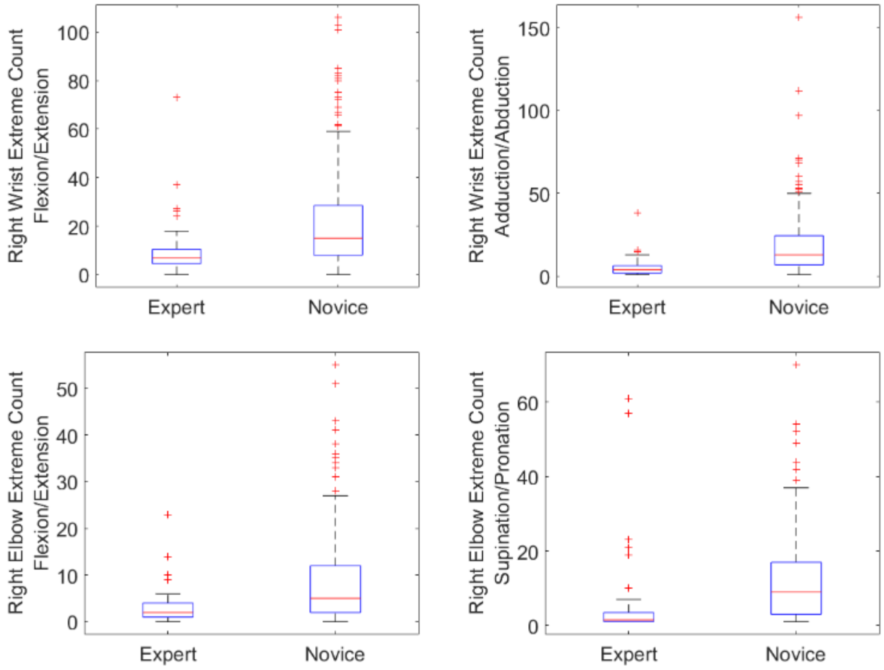

Fig. 7. Number of times an extreme range of motion is entered for novices and experts for right wrist flexion/extension (top-left), right wrist adduction/abduction (top-right), right elbow flexion/extension (bottom-left) and right elbow supination/pronation (bottom-right).

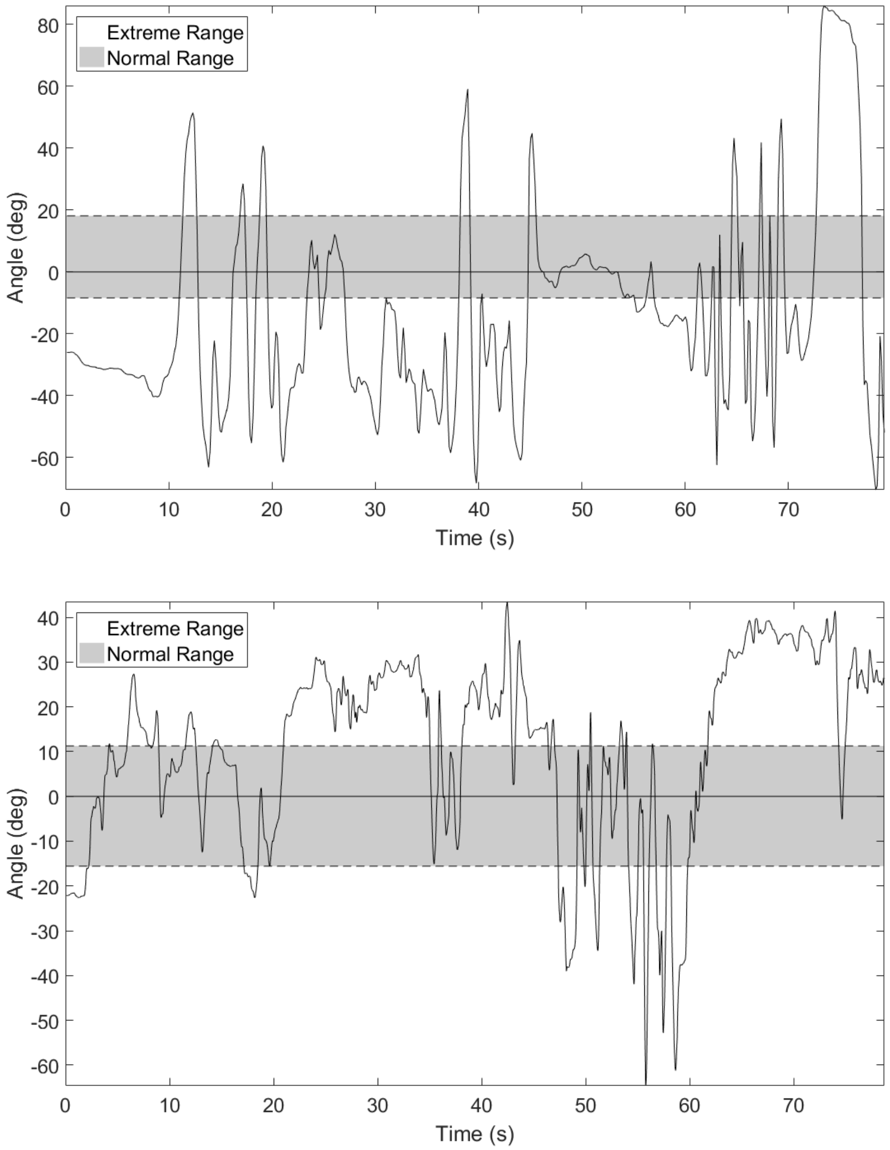

Fig. 8. Example right elbow supination/pronation angle time series for an expert (top) and a novice (bottom) performing a simulated colonoscopy on the same sequence. Grey indicates a normal range of motion.

Table 1. Objective proficiency measures for novice and experts groups. Reported values are medians (inter-quartile range). p indicates the p-value for the Mann-Whitney U test; Δ indicates the effect size Cliff's delta. Asterisk indicates a significant difference between the two groups using the Bonferroni corrected alpha value.

Metric	Novice	Expert	p	Δ
Elapsed Time (s)	96.9 (56.9 - 181.0)	29.0 (21.9 - 40.9)	<0.001*	0.79
Total Hand Path Length (cm)	674.5 (394.1 - 1261.6)	336.1 (249.7 - 499.0)	<0.001*	0.49
Total Hand Motions	72.5 (39.0 - 124.5)	18.5 (13.0 - 26.0)	<0.001*	0.82
Right Wrist Extreme Count Flexion/Extension	15.0 (8.0 - 28.5)	7.0 (4.5 - 10.5)	<0.001*	0.52
Right Wrist Extreme Count Adduction/Abduction	13.0 (7.0 - 24.5)	4.0 (2.0 - 6.5)	<0.001*	0.66
Right Elbow Extreme Count Supination/Pronation	9.0 (3.0 - 17.0)	1.5 (1.0 - 3.5)	<0.001*	0.49
Right Elbow Extreme Count Flexion/Extension	5.0 (2.0 - 12.0)	2.0 (1.0 - 4.0)	<0.001*	0.39
Left Wrist Extreme Count Flexion/Extension	11.0 (5.0 - 22.0)	11.0 (5.0 - 18.5)	0.414	0.07
Left Wrist Extreme Count Adduction/Abduction	9.0 (5.0 - 18.5)	5.0 (2.0 - 8.5)	<0.001*	0.38
Left Elbow Extreme Count Flexion/Extension	1.0 (0.0 - 4.0)	2.0 (1.0 - 3.0)	0.041	0.17
Left Elbow Extreme Count Supination/Pronation	1.0 (1.0 - 1.0)	1.0 (1.0 - 1.0)	0.224	0.05
Right Wrist Extreme Time Flexion/Extension (s)	35.9 (18.0 – 81.1)	11.9 (6.7 – 22.0)	<0.001*	0.56
Right Wrist Extreme Time Adduction/Abduction (s)	54.5 (25.6 – 100.2)	16.2 (8.8 – 26.1)	<0.001*	0.68
Right Elbow Extreme Time Supination/Pronation (s)	82.8 (48.2 – 161.4)	27.3 (19.9 – 38.1)	<0.001*	0.72
Right Elbow Extreme Time Flexion/Extension (s)	56.1 (25.0 – 115.0)	12.9 (2.4 – 25.2)	<0.001*	0.68
Left Wrist Extreme Time Flexion/Extension (s)	47.5 (24.5 – 87.5)	20.3 (15.2 – 30.0)	<0.001*	0.49

Left Wrist Extreme Time Adduction/Abduction (s)	49.8 (25.8 – 97.7)	13.8 (7.4 – 24.3)	<0.001*	0.66
Left Elbow Extreme Time Flexion/Extension (s)	10.0 (0.0 – 53.2)	12.7 (6.6 – 19.9)	0.946	0.01
Left Elbow Extreme Time Supination/Pronation (s)	88.6 (52.7 – 163.8)	26.7 (20.7 – 39.7)	<0.001*	0.78
Right Arm Coordination	0.907 (0.722 - 0.950)	0.898 (0.558 - 0.965)	0.862	0.02
Left Arm Coordination	0.963 (0.872 - 0.985)	0.807 (0.624 - 0.950)	<0.001*	0.42
Right Wrist Total Rotation (rad)	32.0 (19.6 - 68.0)	10.7 (7.9 - 16.6)	<0.001*	0.71
Right Elbow Total Rotation (rad)	31.0 (19.4 - 62.8)	9.7 (6.5 - 17.2)	<0.001*	0.67
Left Wrist Total Rotation (rad)	22.3 (12.0 - 48.4)	17.2 (10.8 - 26.2)	0.015	0.21
Left Elbow Total Rotation (rad)	12.2 (6.2 - 25.1)	8.5 (5.5 – 12.4)	0.003	0.25
Right Wrist Rotational Motions	6.0 (1.5 - 14.0)	2.0 (0.5 - 4.0)	<0.001*	0.39
Right Elbow Rotational Motions	6.0 (2.0 - 12.5)	1.0 (0.0 - 4.0)	<0.001*	0.43
Left Wrist Rotational Motions	4.0 (2.0 - 11.0)	5.0 (2.5 - 10.0)	0.869	0.01
Left Elbow Rotational Motions	0.0 (0.0 - 1.0)	0.0 (0.0 - 2.0)	0.608	0.04