



# HHS Public Access

Author manuscript

*Endoscopy*. Author manuscript; available in PMC 2016 September 12.

Published in final edited form as:

*Endoscopy*. 2014 September ; 46(9): 735–744. doi:10.1055/s-0034-1365463.

## Development and initial validation of an endoscopic part-task training box

Christopher C. Thompson<sup>1,2</sup>, Pichamol Jirapinyo<sup>1,2</sup>, Nitin Kumar<sup>1</sup>, Amy Ou<sup>1</sup>, Andrew Camacho<sup>1</sup>, Balazs Lengyel<sup>3</sup>, and Michele B. Ryan<sup>1</sup>

<sup>1</sup>Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts, USA

<sup>2</sup>Harvard Medical School, Boston, Massachusetts, USA

<sup>3</sup>Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts, USA

### Abstract

**Background and study aims**—There is currently no objective and validated methodology available to assess the progress of endoscopy trainees or to determine when technical competence has been achieved. The aims of the current study were to develop an endoscopic part-task simulator and to assess scoring system validity.

**Methods**—Fundamental endoscopic skills were determined via kinematic analysis, literature review, and expert interviews. Simulator prototypes and scoring systems were developed to reflect these skills. Validity evidence for content, internal structure, and response process was evaluated.

**Results**—The final training box consisted of five modules (knob control, torque, retroflexion, polypectomy, and navigation and loop reduction). A total of 5 minutes were permitted per module with extra points for early completion. Content validity index (CVI)-realism was 0.88, CVI-relevance was 1.00, and CVI-representativeness was 0.88, giving a composite CVI of 0.92.

Overall, 82% of participants considered the simulator to be capable of differentiating between ability levels, and 93% thought the simulator should be used to assess ability prior to performing procedures in patients. Inter-item assessment revealed correlations from 0.67 to 0.93, suggesting that tasks were sufficiently correlated to assess the same underlying construct, with each task remaining independent. Each module represented 16.0%–26.1% of the total score, suggesting that no module contributed disproportionately to the composite score. Average box scores were 272.6 and 284.4 ( $P=0.94$ ) when performed sequentially, and average score for all participants with proctor 1 was 297.6 and 308.1 with proctor 2 ( $P=0.94$ ), suggesting reproducibility and minimal error associated with test administration.

**Conclusion**—A part-task training box and scoring system were developed to assess fundamental endoscopic skills, and validity evidence regarding content, internal structure, and response process was demonstrated.

---

**Corresponding author:** Christopher C. Thompson, MD, MSc, Division of Gastroenterology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA, Fax: +1-617-264-6342, cthompson@hms.harvard.edu.

**Competing interests:** None

## Introduction

Approximately 70 million Americans are affected by digestive diseases, and 15–20 million endoscopic procedures are performed annually in the USA [1,2]. Notwithstanding their clinical utility and beneficial effect on patient care, endoscopic procedures currently carry significant complication rates: 0.95% for screening colonoscopy, 1.07% for upper endoscopy, and as high as 15% in selected therapeutic endoscopy procedures, such as endoscopic retrograde cholangiopancreatography [3–5]. Despite the common nature of these procedures, there is no standard methodology for the training of endoscopists, or for the objective determination of when a trainee has met important thresholds or attained technical competence. Furthermore, there is no objective tool to determine when a trainee is sufficiently familiar with the endoscope and endoscopic equipment to begin clinical procedures [6].

Over the past decade, there has been a steady increase in the development and use of simulators to train and assess endoscopic skills. Limited data on virtual reality simulators have demonstrated some benefits in early trainees [7–11]. However, the high cost associated with computer models has limited their widespread use. In contrast, using simulators as assessment tools has been more controversial. Possible uses for simulators in endoscopy include skill development and evaluation prior to initial human cases, assessment for trainee advancement, and possibly providing an objective method of endoscopist certification and credentialing.

Given the recognized need for simulation in endoscopic training, the development and validation of training and assessment tools are essential. Validity is defined as “appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores,” and validation is defined as the hypothesis-driven “process of accumulating evidence to support such inferences” [12]. Within this contemporary framework, validity integrates evidence from multiple sources including test content, internal structure, response process, relationships to other variables, and consequences of testing. This is in contrast to the previous framework, in which validity was subdivided into “types of validity” (face, content, construct, and criterion-related), and which is no longer used following the most recent consensus standards of the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, published in 1999 [13, 14].

In this study, fundamental endoscopic skills were identified, and the Thompson Endoscopic Skills Trainer (TEST) was developed with the aim of improving and assessing the skills required to perform endoscopy. This report details simulator development and evaluates the validity of the scoring system.

## Materials and methods

The methods consisted of a two-stage process. First, in the development stage, the technical skills that are fundamental to the performance of endoscopy were identified, and part-task prototype modules were developed to simulate those skills. A scoring system was also

iteratively created to quantitatively describe subject performance on the simulation test. Second, the validation stage involved assessment of validity and included expert panel review of each part-task module. In addition, participants with known differences in ability and experience were tested on each module and the scoring system was evaluated.

### Stage 1–Development stage

**Identification of skills to be assessed**—Our previous work in kinematic motion analysis to deconstruct the endoscopic examination into essential maneuvers was reviewed in order to generate a list of maneuvers. A literature review by MEDLINE search was also performed, and 10 experts were interviewed to develop two additional lists of essential endoscopic skills. A panel of three additional experts was then employed to categorize and rate the skills for overlap and relative importance, and a final list of fundamental skills was generated.

**Prototype development**—Using the final list of fundamental skills, several part-task modules were developed to mimic each skill. Prototypes were constructed from polystyrene, cardboard, and plastic to allow rapid modification. Each prototype design for each given skill was then scored for consistency, and the highest ranked design was selected for incorporation into the final training box. Through the development process, the prototypes were then refined until the expert panel agreed that each module accurately reflected the skill of interest.

**Scoring systems and prototype refinement**—Multiple scoring systems were developed, and each scoring system was evaluated. Nine participants with variable endoscopic skills underwent training box evaluation using each scoring system. Three participants were interventional endoscopy attending physicians, three were gastroenterology fellows (one from each training year), and three were novices. Participants were asked to complete each module three times. The scoring system that demonstrated reproducibility (consistent scores upon repetition) and generalizability (endoscopists with a similar level of experience scored within the same score range) was selected.

Following the selection of the scoring system, the prototype was further refined based on feedback and observation in an effort to achieve scoring differentials between experience levels.

### Stage 2–Validation stage

**Definition of validity**—Validity of an assessment tool integrates five main sources of evidence, including content, internal structure, response process, relationships to other variables, and consequences of testing. In the current study, the first three sources of evidence were provided to support validity of the endoscopic part-task training box and its scoring system.

1. Content evidence demonstrates consistency between instructional objectives and the items in the assessment tool, typically in the form of expert agreement. In the current study, a panel of expert endoscopists was

asked to rank each endoscopic part-task for its realism, relevance, and representativeness.

2. Internal structure refers to statistical properties of examination questions or performance prompts that assess level of difficulty and effect on total test scale, and evaluate reproducibility. In the current study, statistical analysis was applied to assess the effect of individual module scores on final composite scores, inter-item correlation, and the reproducibility of test performance.
3. Response process refers to verification of data integrity regarding test administration. In the current study, this involved assessing the effects of simulator administration on training box scores [14, 15].

**Content evidence**—A group of eight experts independently rated each part-task module on a four-point scale for realism (whether the maneuvers required in each part-task were realistic), relevance (whether the skills required in each part-task were relevant for endoscopic skill assessment), and representativeness (whether the maneuvers required in each part-task adequately tested the designated skill). Scores for realism were assessed using basic summary statistics. Content validity index (CVI) for relevance and representativeness was calculated for each module (the proportion of experts who rate the item as content valid defined as a rating of 3 or 4) and for the entire simulator (computed by averaging the item CVI across modules).

In addition, participants with variable endoscopic experience who attended the Annual New York Course of the New York Society Gastrointestinal Endoscopy in December 2011 and Digestive Disease Week in May 2012 were asked to comment on the simulator prototype (Table 1). Specifically, all participants completed each module prior to evaluating the ability of the part-tasks to differentiate endoscopic skill level.

**Internal structure**—The same nine participants who took part in the training box development stage were asked to complete all modules after the prototype was finalized. Inter-item correlation was calculated in order to evaluate the internal structure of the simulator, and measured whether tasks assessed the same underlying construct without significant overlap with other tasks. In addition, the percentage contribution of each module score to the composite training box score was calculated in order to ensure balance and dimensionality. Finally, all participants were asked to consecutively repeat the training box with the same assistant. After a leading period of three tests to eliminate initial learning curve, scores were reviewed to assess score reproducibility.

**Response process**—A group of three experts drafted and revised specific and detailed instructions to ensure clarity and uniformity, before these were given to participants. In addition, written instructions regarding training box test administration were prepared and reviewed with all proctors to encourage standardized endoscopic assistance and test administration. Nine participants were asked to complete the box simulation two additional times. Impact of test administration on score was then assessed by assigning different proctors to sequentially administer the simulation to the same participants. This process was

randomized such that some participants were initially evaluated with proctor 1 and others with proctor 2. Scores were compared to ensure minimal error associated with test administration.

### Statistical analysis

Statistics are reported as mean  $\pm$  SEM. Means were compared using Student's *t* test. CVI was calculated using the proportion of items on an instrument that achieved a rating of 3 or 4 on a scale of 1–4 by content experts [16]. Correlations were calculated using Pearson's correlation coefficients. All statistics were performed using Stata 11.0 (StataCorp, College Station, Texas, USA).

## Results

### Stage 1–Development stage

**Identification of skills to be assessed**—Our group previously performed kinematic analytic studies, which identified important differences in technique and gestures between novice and expert endoscopists [17, 18]. In these studies, the colonoscopic examination was deconstructed into basic elements using advanced kinematic motion analysis. This was done by applying a set of electromagnetic sensors to the endoscope during colonoscopy on a standard latex colon model and during human colonoscopies. These studies revealed that experts required less time, had reduced path length, decreased flex, less tip angulation, reduced absolute roll, and lower endoscope curvature. When compared with novices, the experts were more efficient at scope advancement and performing torque, tip deflection, and loop reduction (Fig. 1).

Review of professional society recommendations and published literature led to the generation of a list of essential endoscopic skills [6, 19–22]. The core skills were divided into two categories (cognitive and technical), and the technical skills were considered for inclusion in the training box (Table 2).

The technical skills identified by practicing endoscopists as being essential to endoscopy were (in descending frequency) loop reduction, endoscope advancement including navigation of turns, snare polypectomy, retroflexion, use of control knobs for orientation and targeted biopsy, controlled endoscope withdrawal, use of accessory devices while maintaining endoscope control and stable endoscope position, keeping the lens clean, use of suction and irrigation, gauging force application, taking pictures at appropriate times, and responding to patient discomfort.

Following expert review of the kinematic analysis, literature review, and endoscopist opinion the fundamental skills that were most consistently identified included loop reduction, accurate use of control knobs, use of torque, goal-directed navigation including turns, achieving and performing tasks in retroflexion, snare polypectomy, and use of forceps for targeted biopsy or other maneuvers.

**Prototype development**—More than 20 part-task prototypes were designed to simulate the skills identified. Of these, nine showed significant promise and were developed further (Table 3).

The endoscopic skill part-tasks were selected for inclusion into the training box based on the degree to which the task reflected fundamental skills, similarity to in vivo endoscopic maneuvers, engineering practicality, durability, and cost. In addition, in order to maintain an acceptable examination time and number of modules, it was necessary to incorporate more than one technical skill into certain part-tasks. Navigation and loop reduction were combined into one module. Use of forceps (as in foreign body removal or targeted biopsy) was also included in four of the modules.

The five part-task modules were then incorporated into a formal training box prototype (Fig. 2). Multiple revisions were then required to assure that all exercises could be accomplished without bias due to gravity or lighting. Specifically, box dimensions were altered to prevent the endoscope from drifting out of position due to gravity. Wall color was adjusted until good lighting with limited glare was obtained. The shape of module entrances was modified to allow better scope maneuverability and stability. In addition, multiple endoscopic accessories were tested to determine which were optimal for maneuvering the various elements in each part-task.

**Scoring system**—Multiple scoring systems were developed and tested to account for both precision and speed. Initial scoring systems required completion of all tasks within each module, with assessment of time to completion and errors (such as dropped items). This led to variable examination times, which were not practical for less-skilled participants. Subsequent scoring systems limited the available time per module. During the design of the box and the scoring system, the mean score for each module was normalized. The difficulty level of each module was iteratively modified and retested. Furthermore, differentiation between novices, fellows, and attending physicians was reassessed and maintained.

The final scoring system allowed 5 minutes for each part-task module. In four modules (knob control, torque, retroflexion, and polypectomy), 10 objects had to be endoscopically maneuvered. For each object that was successfully maneuvered, 10 points were awarded. In the fifth module (loop reduction and navigation), five objects had to be maneuvered. In this test, for each object that was successfully maneuvered, 20 points were awarded. In order to reward efficiency of movement, one point was awarded for each second remaining after task completion. Each module was equally weighted in the total box score. To prevent the overall score from being disproportionately affected by a single module, the upper limit of any individual module was adjusted to twice the value of the second highest module score.

**Prototype refinement**—Following selection of the final scoring system, the prototype was further refined. Nine participants with variable endoscopic experience were recruited to perform each part-task module several times (Table 4). After each trial, participant observations and scores were collected. Each module was modified based on feedback and independent observation of performance, in an effort to achieve scoring differentials between experience levels. Final adjustments are outlined in Table 5.

The beta prototype of the endoscopic part-task training box was developed from the final design of the alpha prototype. The beta prototype had the same dimensions as the final alpha prototype, but was made from white acrylic materials (Fig. 3).

## Stage 2—Validation stage

**Content evidence**—Assessment of the training box and attendant scoring system content by a separate group of eight experts yielded high scores using CVI (Table 6). CVI for realism was 0.88 for all tasks, signifying that experts considered the skills assessed by training box tasks to be realistic compared with clinical endoscopy. The CVI for relevance was 1.00 in all tasks, indicating that experts universally considered each skill relevant to endoscopy. The CVI for representativeness, or coverage of all aspects of each specific skill, was 0.88 for all modules. In addition, the training box was rated to capture all fundamental endoscopy skills with a CVI of 0.88. Composite CVI for the training box was 0.92. The reviewers were in substantial agreement that the training box had demonstrated content-related validity, specifically relevance, representativeness, and realism.

In addition, 54 participants with variable endoscopic experience completed the questionnaire regarding their impression of the simulator. Overall, 82% of participants thought that the training box was able to differentiate between levels of endoscopic experience, and 93% thought that the training box should be used as a practice tool prior to initiation of clinical endoscopy. The opinions were not significantly different between boarded endoscopists (private practitioners, attending physicians, interventional attending physicians) and trainees (fellows, novices) (Table 7).

**Internal structure**—The inter-item consistency of the internal structure was assessed with a correlation matrix of individual module scores (Table 8). Correlations ranged from 0.67 to 0.93, suggesting that tasks were sufficiently correlated to assess the same underlying construct, but that each task was independent and distinct from other tasks. The percentage contribution of each module score to the composite training box score was calculated to assess balance and dimensionality. Each module contributed between 16.0% and 26.1% of the total score. Outperformance on a single module did not result in a high overall score (Table 9).

Participants used the training box several times with the same proctor. After a lead-in period of three tests to account for a foreseen learning curve, the data from two consecutive tests were included. Average total box scores were 272.6 and 284.4 ( $P=0.94$ ) when performed consecutively. This finding supports the other results regarding internal structure of the training box. Specifically, the finding that two consecutive uses of the training box by the same participant with the same proctor resulted in highly similar scores attests to the validity of the internal structure.

**Response process**—Test performance was not affected by proctor performance. The mean score for all participants using proctor 1 was 297.6 and 308.1 using proctor 2 ( $P=0.94$ ). In addition, the printed instructions were successfully used by proctors in all cases. The finding that scores were highly similar when the same participant used two different



proctors demonstrates the validity of the response process, and specifically, the test administration.

## Discussion

Effective development of endoscopic proficiency and objective assessment of endoscopic ability pose significant challenges [23]. In the past, procedure numbers were used as a surrogate for technical proficiency. However, the required number of procedures to achieve competence remains unknown and does not correlate with high-quality performance [24, 25]. Focus then shifted to general competence assessments, including tools such as the validated Mayo Colonoscopy Skills Assessment Tool (MCSAT), which are often cumbersome and subjective but have proven value [19]. Over the past decade, multiple simulators have also been developed. However, none have been independently validated as an endoscopic skill assessment tool [15, 26]. In addition, these simulators have substantial limitations. Tissue-based simulators require special animal use endoscopes, the procurement of animal organs, and extensive preparation and disposal processes. Computerized virtual reality simulators are often cost prohibitive and only allow performance and assessment of entire procedures, rather than part-tasks. In addition, they are often inconveniently located away from areas of clinical activity, which limits access. Other mechanical simulators have the benefit of being inexpensive and convenient; however, they also tend to evaluate entire procedures and have been criticized for lacking realism and applicability [27].

Using contemporary methods of content development, the TEST box was developed by our group to specifically assess core endoscopic skills and to serve as a tool to allow trainees to practice fundamental endoscopic maneuvers prior to initiation of clinical cases. This study presented validity evidence to support the use of the training box and attendant scoring system for the assessment of endoscopic skills.

This study provides strong content-related validity evidence by verifying that each part-task module represents a fundamental endoscopic skill necessary for competence in endoscopy. The rigorous development strategy employed kinematic analysis of endoscopic procedures, literature review, and expert opinions. These were reviewed and prioritized by a panel of experts, and five part-task modules were then refined to reflect the relevant core skills. The training box was then judged by another set of eight experts in the field who deemed that each module was relevant (CVI for relevance of 1.00) and adequately represented fundamental endoscopic skills (CVI for representativeness of 0.88). This makes construct under-representation and construct irrelevance, the two major threats to validity, less likely [28]. In addition, a separate group of 54 participants with variable endoscopic experience also agreed that it was able to evaluate endoscopic ability (82% agreement). This content-related evidence supports the use of the training box and its scoring system as an endoscopic skill assessment tool.

Internal structure validity is suggested by internal consistency among modules. In the current study, assessment of internal structure using inter-item correlation revealed inter-module correlations ranging from 0.67 to 0.93. This range suggests that each task is independent and distinct from other tasks, while still assessing the same underlying



construct: endoscopic skill. As the training box is separated into distinct modules, each primarily targeting one dimension of endoscopic skill, particular skill areas can be assessed independently. This property increases the value of the system as a tool for training and for assessment. In addition, no task contributed a disproportionate amount to the total score, confirming balance and dimensionality. This makes it less likely that outperformance in one module will substantially affect the composite score. Finally, the test appeared to be reliable, as reproducibility was evident when participants were asked to repeat the simulator consecutively.

Response process validity evidence was also demonstrated. Instructions for endoscopists performing the training box were written and revised by a group of experts for accuracy and clarity. They were uniformly read to all participants prior to training box administration. In addition, all proctors were trained to administer the training box in a uniform, unbiased fashion. Similar scores were obtained when each participant was assisted by two different proctors, thereby suggesting high quality control and minimal test administration error.

This study has a few limitations. First, the use of forceps was required in several tasks. It could be conjectured that this would affect the assessment of the targeted skill. However, isolated proficiency in using an accessory tool would not result in a good score in any particular module. Second, the training box does not include all skills that were listed by expert opinion, such as insufflation and suction. Inclusion of these maneuvers would be challenging in a mechanical simulator model. Moreover, CVI for representativeness from the study was 0.88, suggesting that experts agreed that the final prototype adequately represented all fundamental endoscopic skills. Finally, it is important to note that this simulator was not developed to address the cognitive elements of endoscopic training, which are at least as important as the technical elements that were addressed. In addition, this simulator is not meant to be an all-encompassing evaluation system, but should be used in conjunction with appropriate cognitive assessments and the traditional, more subjective, appraisal of endoscopic technique that is currently relied upon.

In the future, we believe that in addition to assessing endoscopic skills, this tool will effectively accelerate and improve endoscopic training. First, the training box allows identification of specific technical deficiencies that would benefit from targeted instruction or practice and measured repetition in those specific areas. Second, this training system is anticipated to be affordable and easily accessible. It can be located in the endoscopy unit and used with a standard endoscope. Third, a full assessment can be completed in 30 minutes, which should allow most clinical fellows with a busy schedule to use the simulator regularly. Finally, the questionnaire supports the use of the simulator as a training tool prior to endoscopy in patients (93% agreement).

Moreover, this training box may be an important component in the evaluation of trainees by providing an objective assessment of technical endoscopic proficiency or a means of charting progress. It may also provide objective milestones for specific levels of training, such as minimal thresholds for advancing to first human cases, or for proceeding from the first to the second year of training. Furthermore, this may be useful beyond the assessment

of trainees, and may play a role in the credentialing and re-credentialing process for practicing endoscopists.

In conclusion, this study applied established methods to develop and validate an endoscopic skills assessment instrument. A larger multicenter study to demonstrate the correlation between training box scores and prior endoscopic experience (relationship to other variables validity evidence) is now required. In addition, a study to assess consequences validity evidence, in which the skills developed in the training box predict improvement in clinical performance, is also warranted.

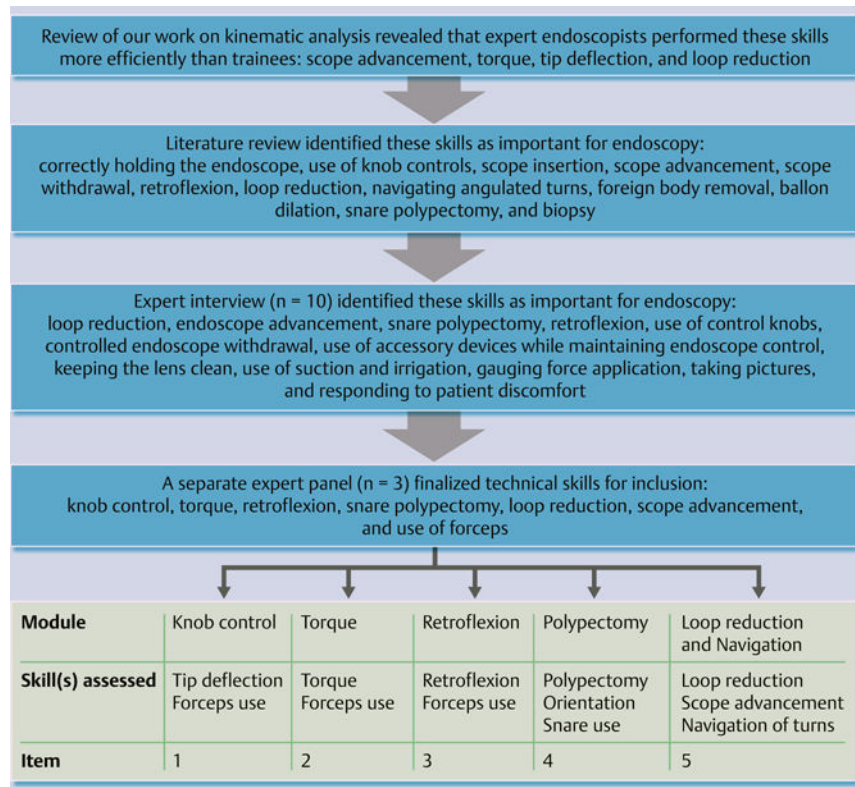
## Acknowledgments

Materials and financial support was provided by the Department of Medicine, Brigham and Women's Hospital, the Wyss Institute for Biologically Inspired Engineering, Harvard Digestive Diseases Center at Harvard Medical School (DK034854), the Center for Integration of Medicine and Innovative Technology (CIMIT), and the National Science Foundation Graduate Research Fellowship Program. We would also like to thank Dr. Daniel Koretz for his assistance with the manuscript revision.

## References

1. National Institutes of Health, U.S. Department of Health and Human Services. Opportunities and Challenges in Digestive Diseases Research: Recommendations of the National Commission on Digestive Diseases. Mar.2009 NIH Publication No. 08-6514
2. Cherry, DK.; Hing, E.; Woodwell, DA.; Rechsteiner, EA. National Ambulatory Medical Survey: 2006 Summary. 3. Hyattsville: National Center for Health Statistics; 2008.
3. Leffler DA, Kheraj R, Garud S, et al. The incidence and cost of unexpected hospital use after scheduled outpatient endoscopy. *Arch Intern Med.* 2010; 170:1752–1757. [PubMed: 20975024]
4. Mallery JS, Baron TH, Dominitz JA, et al. Standards of Practice Committee. American Society for Gastrointestinal Endoscopy. Complications of ERCP. *Gastrointest Endosc.* 2003; 57:633–638. [PubMed: 12709688]
5. Freeman ML, Nelson DB, Sherman S, et al. Complications of endoscopic biliary sphincterotomy. *N Engl J Med.* 1996; 335:909–918. [PubMed: 8782497]
6. Sedlack RE, Shami VM, Adler DG, et al. Colonoscopy core curriculum. *Gastrointest Endosc.* 2012; 76:482–490. [PubMed: 22898404]
7. Cohen J, Cohen SA, Vora KC, et al. Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. *Gastrointest Endosc.* 2006; 64:361–368. [PubMed: 16923483]
8. Sedlack RE, Kolars JC. Computer simulator training enhances the competence of gastroenterology fellows at colonoscopy results of a pilot study. *Am J Gastroenterol.* 2004; 99:33–37. [PubMed: 14687137]
9. Gerson LB. Evidence-based assessment of endoscopic simulators for training. *Gastrointest Endosc Clin N Am.* 2006; 16:489–509. vii–viii. [PubMed: 16876721]
10. Haycock AV, Youd P, Bassett P, et al. Simulator training improves practical skills in therapeutic GI endoscopy: results from a randomized, blinded, controlled study. *Gastrointest Endosc.* 2009; 70:835–845. [PubMed: 19559433]
11. Shirai Y, Yoshida T, Shiraiishi R, et al. Prospective randomized study on the use of a computer-based endoscopic simulator for training in esophagogastroduodenoscopy. *J Gastroenterol Hepatol.* 2008; 23:1046–1050. [PubMed: 18554236]
12. Joint Committee on Standards for Educational and Psychological Testing APA American Educational Research Association and National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: AERA Publications; 1985.
13. Cureton, EE. Validity. In: Lindquist, EF., editor. Educational measurement. Washington, DC: American Council on Education; 1951. p. 621-694.

14. Joint Committee on Standards for educational and Psychological testing APA American Educational Research Association and National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: 1999.
15. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ.* 2003; 37:830–837. [PubMed: 14506816]
16. Grant JS, Davis LT. Selection and use of content experts in instrument development. *Res Nurs Health.* 1997; 20:269–274. [PubMed: 9179180]
17. Spofford IS, Kumar N, Obstein KL, et al. Deconstructing the colonoscopic examination: preliminary results comparing expert and novice kinematic profiles in screening colonoscopy. *Gastrointest Endosc.* 2011; 73:AB415–416.
18. Obstein KL, Patil VD, Jayender J, et al. Evaluation of colonoscopy technical skill levels by use of an objective kinematic-based system. *Gastrointest Endosc.* 2011; 73:315–321. [PubMed: 21111413]
19. Sedlack RE. The Mayo Colonoscopy Skills Assessment Tool: validation of a unique instrument to assess colonoscopy skills in trainees. *Gastrointest Endosc.* 2010; 72:1125–1133. [PubMed: 21111866]
20. Rex DK, Petrini JL, Baron TH, et al. Quality indicators for colonoscopy. *Gastrointest Endosc.* 2006; 63:16–28. [PubMed: 16377310]
21. Rex DK, Bond JH, Winawer S, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol.* 2002; 97:1296–1308. [PubMed: 12094842]
22. ASGE Committee on Training. Colonoscopy Core Curriculum. Mar. 2001 Available from: <http://www.asge.org/assets/0/71328/71340/4beb70579b0546c281fd2feb21d3b144.pdf> Accessed: 20 July 2010
23. Cohen J, Bosworth BP, Chak A, et al. Preservation and incorporation of valuable endoscopic innovations (PIVI) on the use of endoscopy simulators for training and assessing skill. *Gastrointest Endosc.* 2012; 76:471–475. [PubMed: 22809879]
24. Spier BJ, Benson M, Pfau PR, et al. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc.* 2010; 71:319–324. [PubMed: 19647242]
25. Sedlack RE. Training to competency in colonoscopy: assessing and defining competency standards. *Gastrointest Endosc.* 2011; 74:355–366. [PubMed: 21514931]
26. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med.* 2006; 119:166.e7–166.e16. [PubMed: 16443422]
27. Cohen J, Thompson CC. The next generation of endoscopic simulation. *Am J Gastroenterol.* 2013; 108:1036–1039. [PubMed: 23820991]
28. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ.* 2004; 38:327–333. [PubMed: 14996342]



**Fig.1.** Process of identification of core endoscopic skills.



**Fig.2.** Endoscopic part-task training box alpha prototype. **a,b** Knob control, with use of forceps to move objects between compartments. **c,d** Torque, with use of forceps to pick up rings from the center post and place on pegs. **e,f** Retroflexion, with use of forceps to pick up objects in retroflexed position and place on forward wall. **g** Polypectomy, using snare to grasp polyps. **h** Loop reduction, which simulates loop formation and allows a realistic loop reduction maneuver.





**Fig.3.** Endoscopic part-task training box beta prototype. **a** Knob control. **b** Torque. **c** Retroflexion. **d** Polypectomy. **e** Loop reduction and navigation.

**Table 1**

Characteristics of 54 surveyed participants.

	n	Mean years of experience	EGD	Colonoscopy	ERCP or EUS
Interventional	14	13.2	10 458	4709	3609
Private practitioner	11	18.5	8 195	5893	261
Attending physician	8	10.4	2 963	2213	200
Fellow	15	2.8	271	216	18
Novice	6	0.3	42	4	0

EGD, esophagogastroduodenoscopy; ERCP, endoscopic retrograde cholangiopancreatography; EUS, endoscopic ultrasound.



**Table 2**

List of fundamental technical skills important for endoscopy.

– Correctly holding the scope	– Loop reduction
– Use of the scope knob controls	– Angulated turns
– Scope insertion	– Foreign body removal
– Scope advancement (torque)	– Balloon dilation
– Scope withdrawal	– Snare polypectomy
– Retroflexion	– Biopsy

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3**

Nine endoscopic part-tasks were further developed based on kinematic analysis of endoscopic examination, literature review, and expert opinion. Five of these nine part-tasks were selected for inclusion in the final training box.

<b>Core technical skills</b>	<b>Endoscopic part-task exercise</b>	<b>Selected?</b>
Knob control (1)	Drawing a line to connect dots	No
Knob control (2)	Transferring objects from one compartment to another	Yes
Torque (1)	Navigating through cylinders and detaching rings from the walls	No
Torque (2)	Transferring rings from one pole to another	Yes
Retroflexion (1)	Retroflexing to grab hollow cylinders and stacking them onto receptacles	No
Retroflexion (2)	Transferring objects from a front to a back wall	Yes
Snare polypectomy	Snaring an object at its base	Yes
Foreign body removal	Use of forceps to collect differently shaped objects within a maze	No
Loop reduction	Advancing the scope into a preformed loop, then reducing the loop	Yes

**Table 4**

Characteristics of participants who took part in content validation (n = 9).

Level	n	Mean years of experience	EGD	Colonoscopy	ERCP or EUS
Expert: interventional attending physicians	3	10	>4600	>3000	>2500
Intermediate: fellows	3	2	287	213	0
Novice: medical students	3	0	0	0	0

EGD, esophagogastroduodenoscopy; ERCP, endoscopic retrograde cholangiopancreatography; EUS, endoscopic ultrasound.

**Table 5**Parameters adjusted to each part-task. Mean score  $\pm$  SEM.

Adjustment	Expert	Intermediate	Novice
<b>Knob control</b>			
Transfer 10 objects between compartments	135 $\pm$ 16	40 $\pm$ 15	0
Partitioned into 4 compartments; scope insertion depth made constant			
Object materials tested; lubricious materials selected to increase difficulty	171 $\pm$ 81	80 $\pm$ 20	35 $\pm$ 5
Locking mechanism for scope fixation finalized	133 $\pm$ 63	75 $\pm$ 5	20 $\pm$ 6
<b>Torque</b>			
1 ring to be transferred sequentially between posts	139 $\pm$ 53	90 $\pm$ 10	35 $\pm$ 15
5 rings to be transferred from central post on floor to 5 posts on walls			
Posts tilted upward to ease ring placement; height and location of central post adjusted	172 $\pm$ 28	96 $\pm$ 28	43 $\pm$ 7
Object materials tested; rubbery rings selected	141 $\pm$ 53	80 $\pm$ 27	45 $\pm$ 5
Wall posts moved and staggered to force torque movement			
<b>Retroflexion</b>			
10 objects to be transferred from the front to the back wall	30 $\pm$ 5	–	0
10 objects to be transferred from the back to the front wall			
Length adjusted; posts tilted upward to prevent objects from falling; distance between posts adjusted to optimize object placement and full retroflexion	128 $\pm$ 13	35 $\pm$ 15	25 $\pm$ 5
Object materials tested; gummy material selected			
Entrance slot raised; front wall posts adjusted to allow full retroflexion	157 $\pm$ 27	66 $\pm$ 10	25 $\pm$ 5
<b>Polypectomy</b>			
10 round objects to be snared and detached from different walls	229 $\pm$ 34	50 $\pm$ 20	57 $\pm$ 9
Moved some polyps to corners to add challenge	215 $\pm$ 9	133 $\pm$ 53	40 $\pm$ 15
Switched from round to flat objects to encourage snaring polyp base			
Magnet, Velcro, rubber corks, earring backs tested; electronic polyps selected	253 $\pm$ 16	124 $\pm$ 44	35 $\pm$ 5
<b>Navigation<sup>1</sup></b>			
2 different paths to be navigated, with target task at the end	–	30 $\pm$ 5	60 $\pm$ 8
5 paths; rearranged paths to equalize difficulty			
Navigation tubes lined with soft pad and film to protect scope and reduce resistance	163 $\pm$ 45	118 $\pm$ 58	80
Target tasks tested: bell-ringing, tip deflection; foreign body retrieval selected			
Alpha loop reduction task implemented prior to the navigation task	141 $\pm$ 30	60 $\pm$ 12	40 $\pm$ 12
<b>Loop reduction<sup>1</sup></b>			
Tester required to scope through a preformed loop, then reduce			
Pipe materials tested: PVC, flexible pipe, carved polystyrene	141 $\pm$ 30	60 $\pm$ 12	40 $\pm$ 12
Loop shape adjusted to grasp scope; transparent roof implemented to support scope during loop reduction; alpha loop reduction moved before navigation			

PVC, poly(vinyl chloride)

<sup>1</sup>Loop reduction and navigation tasks were combined into a single part-task, reported as duplicate test scores.

**Table 6**

Content validity index.

	<b>CVI</b>
<b>Polypectomy</b>	
Realism	0.88
Relevance	1.00
Representativeness	0.88
<b>Retroflexion</b>	
Realism	0.88
Relevance	1.00
Representativeness	0.88
<b>Torque</b>	
Realism	0.88
Relevance	1.00
Representativeness	0.88
<b>Tip deflection</b>	
Realism	0.88
Relevance	1.00
Representativeness	0.88
<b>Navigation/Loop reduction</b>	
Realism	0.88
Relevance	1.00
Representativeness	0.88
Does the training box include all core skills?	0.88
Composite CVI	0.92

CVI, content validity index.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 7**

Impression of the training box by surveyed participants.

	<b>Experts</b>	<b>Trainees</b>	<b>P value</b>
The training box is able to differentiate level of endoscopic experience, %	Agree 83 Disagree 17 No opinion 0	Agree 80 Disagree 20 No opinion 0	0.94
The training box should be used prior to the first human case, %	Agree 97 Disagree 3 No opinion 0	Agree 86 Disagree 7 No opinion 7	0.22

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 8**

Inter-item consistency, module score.

	<b>Knob control</b>	<b>Torque</b>	<b>Retroflexion</b>	<b>Polypectomy</b>	<b>Loop and navigation</b>
Knob control	1.000	0.712	0.837	0.861	0.926
Torque	0.929	1.000	0.903	0.668	0.690
Retroflexion	0.866	0.850	1.000	0.850	0.866
Polypectomy	0.690	0.668	0.903	1.000	0.929
Loop/navigation	0.926	0.861	0.837	0.712	1.000



**Table 9**

Percentage of points attributable to each module.

<b>Part-task</b>	<b>Mean score</b>	<b>SEM</b>	<b>Percentage of total</b>
Knob control	76.1	18.1	16.0
Torque	99.2	13.3	20.8
Retroflexion	79.1	16.0	16.6
Polypectomy	124.6	19.4	26.1
Loop and navigation	97.7	15.4	20.5
Total score	484	37.6	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript