Automatic and unbiased assessment of competence in colonoscopy

exploring validity of the Colonoscopy Progression Score (CoPS)

Preisler, Louise; Svendsen, Morten Bo Søndergaard; Søndergaard, Bo; Brink, Lene; Nordentoft, Tyge; Svendsen, Lars Bo; Konge, Lars

Automatic and unbiased assessment of competence in colonoscopy: exploring validity of the Colonoscopy Progression Score (CoPS)

Authors Louise Preisler[1], Morten Bo Søndergaard Svendsen[2,3], Bo Søndergaard[4], Lene Brink[5], Tyge Nordentoft[5], Lars Bo Svendsen[1], Lars Konge[2]

Institutions Institutions are listed at the end of article.

Corresponding author
Louise Preisler, MD
Rigshospitalet – Surgical
Department C, 2122
Blegdamsvej 7 – 9
Copenhagen 2100
Denmark
Fax: +45-35458618
louise.preisler.01@regionh.dk

Background and aims: Colonoscopy is a difficult procedure to master. Increasing demands for colonoscopy, due to screening and surveillance programs, have highlighted the need for competent performers. Valid methods for assessing technical skills are pivotal for training and assessment. This study is the first clinical descriptive report of a novel colonoscopy assessment tool based on Magnetic Endoscopic Imaging (MEI) data and the aim was to gather validity evidence based on the data collected using the "Colonoscopy Progression Score" (CoPS).

Methods: We recorded 137 colonoscopy procedures performed by 31 endoscopists at three university hospitals. The participants performed more than two procedures each (range 2 – 12) and had an experience of 0 – 10 000 colonoscopies. The CoPS was calculated for each recording and validity was explored using a widely accepted contemporary framework. The following sources of validity evidence were explored: response process (data collection), internal structure (reliability), relationship to other variables (i.e. operator experience), and consequences of testing (pass/fail).

Results: Identical set-ups at all three locations ensured uniform data collection. The Generalizability coefficient (G-coefficient) was 0.80, and a Decision-study (D-study) revealed that four recordings were sufficient to ensure a G-coefficient above 0.80. We showed a positive correlation between CoPS and experience with Pearson's r of 0.61 ($P < 0.001$). A pass/fail standard of 107 points was established using the contrasting group method to explore the consequences of testing.

Conclusion: This study provides evidence supporting the validity of the CoPS for use in assessing technical colonoscopy performance in the clinical setting.
Study registration: NCT01997177.

Background

▼

The colonoscopy procedure is challenging to learn. It requires a high level of hand-eye coordination, dexterity, and a broad variety of maneuvers. The variability in shape and length of the colon as well as the tendency for loop-formation require considerable experience [1]. Screening and surveillance programs have highlighted the importance of skilled endoscopists, as colonoscopy is a procedure with rare but serious complications for the patient, correlated to inexperience of the endoscopists [2]. Timely certification in competency-based training programs requires a reliable and valid assessment method.

Several studies have tried to determine the number of procedures needed to reach competency in colonoscopy. Some authors have suggested 140 or 275 procedures; others more than 500 procedures [3 – 5]. The span in numbers indicates that endoscopists acquire technical skills at different rates and the number of procedures alone cannot be used as the sole parameter in order to ensure competency.

Intubation of the cecum is mandatory in order to examine the entire colon. Because of this, cecal intubation rate is often reported as a measure of competency [6]. Another parameter often used to measure competency in colonoscopy is adenoma detection rate (ADR). Whether these single parameters used alone or in combination can be used to measure competency is questionable. A study on reliability using ADR revealed that a very large sample size (e.g. 500 procedures) was needed in order to ensure reliable assessment [7]. Existing tools assessing process-orientated technical skills are rater dependent [5,8,9]. This in turn makes them costly, time demanding, and prone to bias as they are based upon expert observations.

It has been suggested that virtual reality simulators could be used for certification [10]. However,
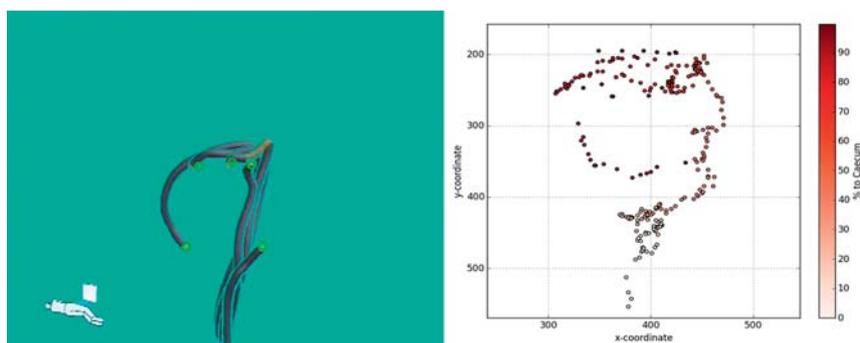
Fig. 1 Colonoscopy progression score acquisition from the MEI unit in a virtual grid. The left image illustrates five chosen frames from a colonoscopy tracking. The dots on the right image represent time and route to cecum (color coded for visual interpretation).

the simulator models have limited discriminatory ability as measured by composite score to determine competency [11]. Generally, the most technically demanding part of the diagnostic colonoscopy procedure is the advancement of the colonoscope to the cecum. The inexperienced endoscopist tends to go slowly [12], having trouble passing bends and resolving loops [13]. Perforations are rare but are associated with high morbidity and mortality for the patient [14]. Most perforations occur as a result of excessive force applied to the endoscope tip or from stretching of the bowel due to loop formation [15]. In a teaching perspective, it seems obvious to focus on the smooth, safe, and gradual progression of the colonoscope in order to avoid complications.

Magnetic endoscope imaging (MEI) equipment, such as the 3D magnetic Scope guide (Olympus Optical, Tokyo, Japan) visualize in real time the shape of the colonoscope during the procedure. MEI provides an image of the scope but provides no overall feedback. The MEI pattern of progression during the colonoscopy could differentiate between the experienced and the inexperienced performers in colonoscopy [16].

Our aim was to calculate a numerical score to quantify the endoscope's progression in order to create an automated tool to assess advancement of the colonoscope. The "Colonoscopy Progression Score" (CoPS) is a tool we have developed to assess the advancement of the endoscope by tracking it through the colon using recorded MEI sequences and calculating a single numerical score using an algorithm.

In a pilot study, we have shown that the CoPS could discriminate between novices and experts in a simulated environment and could be used for measuring competency [17]. However, there are obvious differences between a standardized simulator set-up and a clinical setting, where procedures are performed on patients.

This study is the first clinical descriptive report of a novel colonoscopy assessment tool and the aim of this study was to assess evidence of validity of the CoPS for use within the clinical setting.

Materials and methods
▼

In an earlier study, we presented the design of the CoPS tool [17]. The CoPS tool is based on MEI technology. MEI has made it possible to visualize the shape of the colonoscope inside the patient during a procedure. Small coils inside the colonoscope generate magnetic impulses, a receiver unit registers the signal from the colonoscope, and a designated computer generates an image of the estimated shape of the colonoscope on a monitor.

We recorded the MEI image and processed the recordings in order to localize the tip of the colonoscope. The recordings were processed using MatLab (MathWorks, Inc., Natick, Mass., United

States). By placing the tip position of the scope into a virtual grid consisting of multiple squares, it was possible to follow the progression of the scope throughout the procedure (○ **Fig. 1**). The CoPS reflects the passage of the colonoscope through the grid as a function of time. The score increases if the tip moves forward from one square to the next and decreases if the tip of the colonoscope move backwards to a square already visited. A smooth progression through the colon and the intubation of the cecum would result in high CoPS, whereas a slow progression would result in lower CoPS. Progress in colonoscopy sometimes involves pulling back in order to change the configuration of the colon. However, this technique will only result in a limited, unavoidable decrease in the score. Trainees do not complete all colonoscopies; sometimes a more experienced supervisor takes over in order to complete the procedure. In case the trainee did not complete the procedure, the CoPS was adjusted accordingly to the end point. We used the following landmarks as end points: cecum (4), hepatic flexure (3), splenic flexure (2), recto-sigmoid (1). We chose these landmarks as they are easy to recognize from the endoscopic image. Only by reaching the cecum could the full CoPS be achieved. Otherwise, the score would be penalized as the obtained CoPS rose to the power of the end point number divided by four. We used the form: $CoPS^{(x/4)}$, x being one of the four end points. This exponential decrease in scores penalized incomplete procedures heavily (using a linear penalty algorithm would result in too high a score for smooth, incomplete procedures).

Setting and participants

The study was conducted at the endoscopy departments of the Copenhagen University hospitals of Herlev, Hvidovre, and Rigshospitalet from 30 November 2013 until 30 June 2014.

At each endoscopy department, a standardized study set-up was installed. The video recorder (UnicDoc, Simonsen & Weel, Denmark) was connected to the MEI and a laptop computer was used to store the recordings. A database (UnicBase, Simonsen & Weel, Denmark) was used to link the recordings to the endoscopists. All endoscopists working in the three units, performing elective diagnostic colonoscopies were included. Participants were asked to fill out a brief questionnaire including demographics and colonoscopy experience (○ **Table 1**). A letter of acceptance of participation was handed out and returned before the study start.

In order to reduce patient variability, only diagnostic procedures on adult patients with no history of colon resection were included in the study. The patients received sedation but procedures under general anesthesia were excluded. Data on level of sedation were not collected.

The study was reported to the Danish ethical committee (H-1-2013-FSP-58). According to Danish law, the study did not require

Experience (No. of prior colonoscopy procedures)	Gender		Specialty	
	Male	Female	Gastroenterology	Surgery
< 50	4 (12.9 %)	6 (19.4 %)	5 (16.1 %)	5 (16.1 %)
50 – 499	3 (9.7 %)	4 (12.9 %)	4 (12.9 %)	3 (9.7 %)
500 – 10 000	11 (35.5 %)	3 (9.7 %)	9 (29.0 %)	5 (16.1 %)

Table 1 Participating endoscopists (n = 31) and distribution of experience and specialty.

the consent of the patients in whom the colonoscopy was being performed. The participating endoscopists were informed and gave their full written consent before the study.

Data collection

We used the MEI Olympus Scope guide (Olympus Optical, Tokyo, Japan) and recorded the route of the colonoscope from the anus to the cecum. Visualization of the ileo-cecal valve and the appendix orifice ensured intubation of the cecum. The operator or supervisor located the landmarks – members of the research group were not involved in determining the end point of the scope.

If the trainee had difficulties and a more experienced endoscopist took over before reaching the cecum, the recording was stopped and the end point noted.

Recordings were logged with a number, which corresponded to a different list containing ID and experience. The quality of the recordings was checked before they were processed in the CoPS computer algorithm. The ID of the participants was blinded to the researcher handling the recordings.

Incomplete recordings and recordings containing noise as a consequence of lost or ambiguous MEI signal, and recordings with lost signal due to position change were excluded.

Exploring validity evidence

We used the contemporary framework of validity by Messick [18] exploring five different sources of validity evidence and further described by Downing [19]. (1) Content relates to the relationship between test content and the construct of interest (i.e. inserting the colonoscope). (2) Response process concerns the quality of the gathered assessment data. (3) Internal structure is about the reliability of the test, and (4) relation to other variable is explored by correlating the new assessment to existing forms of assessing skills. Finally, (5) consequences of testing are explored by looking at the intended and unintended consequences of the test.

Content was explored in the pilot study [17] and the identical set-up and computerized data collection ensured a uniform response process.

Data analysis

For internal structure, we used generalizability theory as described by Brennan [20]. Estimation for variance components was conducted using Henderson's Method 1 procedure (sometimes also known as the "analogous-ANOVA procedure"). By estimating components of variance, a G-coefficient was defined and this was used to explore test reliability and internal structure of the tool. A random-effects model in Stata IC (StataCorp, College Station, Texas, United States) was used for finding the confidence interval for the G-coefficient. A Pearson product-moment correlation coefficient was calculated to explore relationships with other variables. We analyzed the correlation between the numbers of procedures performed by the endoscopist and the CoPS of each procedure. Consequences of testing were explored by setting a pass/fail standard using the contrasting groups method

[21]. This method allows decision makers to move the established cut-score depending on their objective (e.g. minimizing false negatives or minimizing false positives). We chose to use the actual contrasting groups' pass/fail standard to explore the consequences. The consequences of the pass/fail standard were reported as frequencies.

For the purpose of standard setting, the participants were divided into two subgroups in order to use the contrasting groups method: A novice group (experience: 0 – 50 colonoscopy procedures) and an experienced group (experience: > 500 colonoscopy procedures). Grouping was based upon the literature [22, 23].

We used a statistical software package (SPSS Inc. version 20.0, Chicago, Illinois, United States) to perform the statistical analysis.

Results
▼

A total of 31 endoscopists participated in the study, performing a total of 206 colonoscopies, of which 137 (66.5 %) procedures were without technical problems. We recorded between 2 and 12 recordings per participant. The endoscopists differed in prior experience from new trainees to very experienced endoscopists (range 0 – 10 000). For details, see ○ **Table 1**.

We explored internal structure by analyzing test reliability and found a Generalizability (G) coefficient of 0.80. By conducting a D-study, we found that each assessment should contain four CoPS measures to achieve a Generalizability coefficient above 0.80 (○ **Fig. 2**); the 95 % confidence interval for the G coefficient: was 0.736 to 0.910.

We demonstrated a positive correlation between the level of endoscopic experience and CoPS with a Pearson's correlation of 0.61 ($P < 0.001$) (○ **Fig. 3**).

Furthermore, we analyzed the difference in CoPS between the novice and experienced groups. Novices achieved a mean CoPS of 31.4 (SD 49.0) whereas the experienced group achieved a mean CoPS of 197.6 (SD 125.4) ($P < 0.001$).

A pass/fail level was established at 107 points using the contrasting groups method (○ **Fig. 4**). The consequence of the pass/fail standard was that none of the novices passed the test and three out of 14 of the experienced group failed the test; 16 of our participants (including the three experienced endoscopists who failed) performed only two or three procedures.

Discussion
▼

To assess technical skills in colonoscopy, we developed the CoPS system that calculates a numerical score, based on the position of the tip of the colonoscope during insertion. We tested the CoPS in a clinical setting in order to determine objective measures of colonoscopy skills and demonstrated a good correlation between CoPS and experience, which indicates that the score can be used for training. However, further studies with sufficient power are
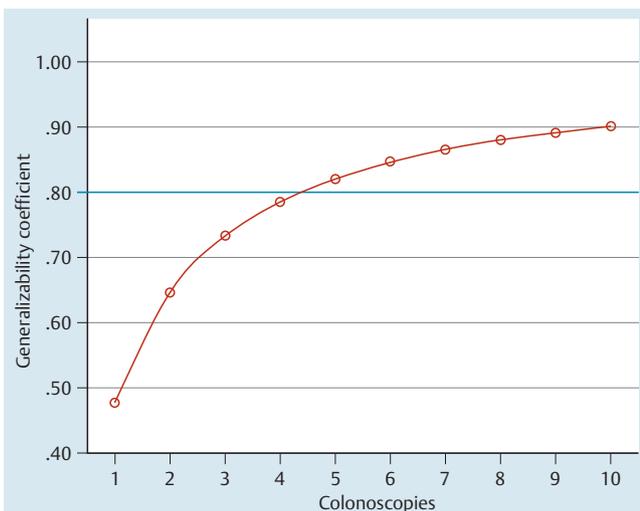
Fig. 2 G-Coefficient for various numbers of colonoscopy procedures; the dotted line indicates the number of procedures needed to reach 0.80.
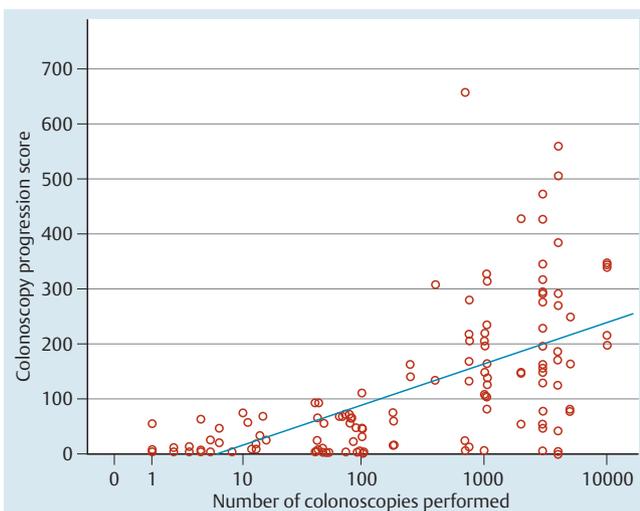
Fig. 3 Performance score according to the level of experience as number of prior procedures. Each dot represents a score correlated to experience.
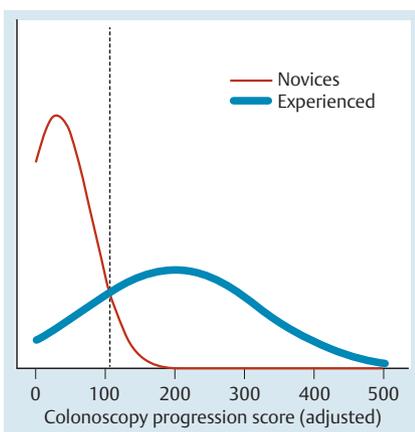
Fig. 4 Standard setting using Contrasting Groups Method. The dotted line indicates pass/fail level.

needed before CoPS can be warranted for certification and re-certification.

We used a contemporary framework for testing for validity evidence, response process, internal structure, relationship to other variables, and consequences of testing [18].

The objective automatic tool and the video-based approach allowed assessment to be blinded and eliminated the potential bias caused by the relationship between the rater and ratee (i.e. subjectivity, false impression, rank).

We found CoPS reliable with a Generalizability coefficient (G-coefficient) of 0.80, and a Decision-study revealed that four recordings were sufficient to ensure a G-coefficient above 0.80 [24]. This makes CoPS a feasible tool to use for assessment of colonoscopy skills.

Existing validated colonoscopy tools such as the American "Mayo Clinic Colonoscopy Assessment Tool" (MSCAT) [25] and the British "Direct Observation of Procedural Skills score" (DOPS) [8] represent global assessment tools assessing motor skills of colonoscopy, as one of several domains.

In a validity study of DOPS, a Generalizability coefficient of 0.81 was reported in a set-up with a sample of two patients and two assessors [8], which implies that four assessments are necessary in order to achieve sufficient reliability. In a study gathering validity evidence of MSCAT, individual motor skills scores with correlation coefficients ranging from 0.59 to 0.83 were reported and the average of these motor scores demonstrated a correlation coefficient of 0.88 [25]. Reliability of both DOPS and MSCAT is equivalent to our findings using CoPS. However, methods relying on rater assessment have disadvantages compared to CoPS, as they are resource intensive and unsuitable for prolonged or repeated performance measurements.

We wanted to explore the relationship to other variables by comparing the CoPS to the external variable: Degree of experience. By using a Pearson's r, we found a correlation of 0.61. Educational correlations in the range of 0.50–0.60 are generally considered to suggest a meaningful correlation with regard to practical value and scores above 0.60 are considered to be substantially correlated [26].

We used the contrasting groups method to set a standard for a pass/fail score and we tested for the consequences of this score. The CoPS tool could identify inexperienced endoscopists as none of the endoscopists in the novice group achieved a mean CoPS above the established pass/fail standard of 107 points. However, three of the experienced endoscopists failed the test. It is noteworthy that the mean scores of all three experienced endoscopists who failed were based on very few observations: two, two, and three colonoscopies, respectively. Even experienced endoscopists can obtain a low CoPS when performing a procedure on a particularly difficult patient. It is important to acknowledge that CoPS was tested in an unselected patient population. We do acknowledge that some colonoscopy cases are very difficult when it comes to insertion [27]; however, in this study, we tested in an unselected patient population and demonstrated a test reliability (G) coefficient > 0.80 measuring four procedures (● **Fig. 2**). We do believe that having four assessments in a test will diminish the impact of difficult cases. In future studies, we would ensure that all participants undertook four procedures in order to use the tool for pass/fail decisions. This is consistent with earlier findings on assessment of endoscopic procedures [28].

Study strength and limitations

The strength of our study was the uniform set-up in three major university endoscopy departments ensuring inclusion of participants with a broad range of experience (0 – 10 000, see ◉ **Fig. 3**). A limitation of the set-up was the challenge to make a technically perfect recording of the MEI signal. We recorded the MEI signal from the monitor but as endoscopists often change the position of the patient or the MEI during the procedure, the MEI signal was not visible at all times on some recordings.

Technical problems with lost signal or noise caused exclusion of a large number of recordings (33.2 %) as it was not possible for the computer to score the CoPS. We have no reason to believe that the excluded procedures differed in any important ways from the included procedures.

In the future, we hope to be able to obtain the 3D-coordinates of the colonoscope directly from the MEI, which would solve the problems related to the video-image grabbing software.

The experience of the participants was self-reported as no database contained these data. This is a limitation of the study, especially for the intermediate and experienced group as novices with little experience are more likely to be accurate in the numbers reported.

Another limitation of the study was that the number of colonoscopy procedures performed by some of the endoscopists was less than four. After conducting a D-study, we realized that sufficient reliability is only reached after a minimum of four assessments. In the future, at least four procedures should be assessed using CoPS to avoid making unjustified decisions based on CoPS.

The CoPS tool assesses only the technical procedure of inserting the colonoscope, which is the hardest task to learn to master in the colonoscopy procedure, however, we do acknowledge that cognitive aspects of the procedure such as ability to interpret findings are also very important in evaluation of competency.

The strength of the CoPS tool is that it does not take extra time or space in the endoscopy suite and provides an immediate objective measure of skills. As stated previously, there are other validated colonoscopy forms assessing technical skills, however, most of these tools rely on rater observations. Using raters for assessment takes time and coordination, which can be hard to manage in endoscopy units with large workloads. CoPS can be used as a continuous assessment throughout training and delivers numerical feedback.

The CoPS provides feedback as a score, and it could be argued that this feedback is too simplistic to provide meaningful feedback of a complex technical maneuver. However, a recent study of feedback challenges previous assumptions suggesting that qualitative comments are superior to a score when it comes to feedback of technical skills. As a supplement to the CoPS, the tool also provides a CoPS map where more specific feedback is provided in a visual form [17]. Ego-orientated feedback in a numerical form (score) was powerful, compared to task-orientated feedback [29]. We found that four recordings were sufficient to ensure reliability, and can be performed on a single day.

In an accreditation era, the perspectives of an accessible and objective tool are obvious. The increasing demand for colonoscopy, due to screening and surveillance programs for colorectal cancer, has highlighted the demand for competent colonoscopists and feasible ways to measure competency. The MEI technology is already available in many endoscopy units and to use the CoPS tool will not create an extra workload or take up the endoscopist's time. It would be interesting to explore the correlation between "Patient discomfort" and CoPS, and we intend to address this in a future study.

The major innovation with regard to the CoPS is the possibility to measure performance in colonoscopy skills in an unbiased way, due to advances in image acquisition, analysis, and high-speed data processing [30].

Conclusion
▼

The need for quantifiable and reproducible measures of skill is fundamental to ensure quality training as well as maintain competency in colonoscopy. We found evidence to support the validity of data collected using a novel tool. The CoPS tool provides an opportunity to assess trainees continuously throughout training in an easy and economical way if the MEI technology is available. In the future, CoPS has the potential of being used to guide training in the endoscopy suite but further studies are required before it can be considered to be a tool to measure competence and ensure maintenance of competency.

Competing interests: None

Institutions
[1] Department of Surgical Gastroenterology and Transplantation, Rigshospitalet, Copenhagen, Denmark and University of Copenhagen, Copenhagen, Denmark
[2] Centre for Clinical Education (CEKU), The Capital Region of Denmark and University of Copenhagen, Copenhagen, Denmark
[3] Marine Biological Sections, Department of Biology, University of Copenhagen, Copenhagen, Denmark
[4] Department of Gastroenterology and Gastrointestinal Surgery, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark
[5] Department of Gastroenterology and Gastrointestinal Surgery, University Hospital Herlev, Herlev, Denmark

References
1 *Saunders BP, Phillips RK, Williams CB*. Intraoperative measurement of colonic anatomy and attachments with relevance to colonoscopy. Br J Surg 1995; 82: 1491 – 1493
2 *Bini EJ, Firoozi B, Choung RJ* et al. Systematic evaluation of complications related to endoscopy in a training setting: A prospective 30-day outcomes study. Gastrointest Endosc 2003; 57: 8 – 16
3 *Spier BJ, Durkin ET, Walker AJ* et al. Surgical resident's training in colonoscopy: numbers, competency, and perceptions. Surg Endosc 2010; 24: 2556 – 2561
4 *Lee S-H, Chung I-K, Kim S-J* et al. An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. Gastrointest Endosc 2008; 67: 683 – 689
5 *Sedlack RE*. Training to competency in colonoscopy: assessing and defining competency standards. Gastrointest Endosc 2011; 74: 355 – 366
6 *Koch AD, Haringsma J, Schoon EJ* et al. Competence measurement during colonoscopy training: the use of self-assessment of performance measures. Am J Gastroenterol 2012; 107: 971 – 975
7 *Do A, Weinberg J, Kakkar A* et al. Reliability of adenoma detection rate is based on procedural volume. Gastrointest Endosc 2013; 77: 376 – 380
8 *Barton JR, Corbett S, Van der Vleuten CP*. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. Gastrointest Endosc 2012; 75: 591 – 597
9 *Walsh CM, Ling SC, Khanna N* et al. Gastrointestinal Endoscopy Competency Assessment Tool: development of a procedure-specific assessment tool for colonoscopy. Gastrointest Endosc 2014; 79: 798 – 807
10 *Sedlack RE, Baron TH, Downing SM* et al. Validation of a colonoscopy simulation model for skills assessment. Am J Gastroenterol 2007; 102: 64 – 74
11 *Walsh CM, Sherlock ME, Ling SC, Carnahan H*. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. In: Walsh CM, ed. The Cochrane Collaboration. Cochrane Database of

Systematic Reviews [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2012 [cited 2 June 2013]

12 *Marshall JB*. Technical proficiency of trainees performing colonoscopy: a learning curve. Gastrointest Endosc 1995; 42: 287 – 291

13 *Haycock A*, *Bassett P*, *Bladen J* et al. Validation of the second-generation Olympus colonoscopy simulator for skills assessment. Endoscopy 2009; 41: 952 – 958

14 *Lüning TH*, *Keemers-Gels ME*, *Barendregt WB* et al. Colonoscopic perforations: a review of 30,366 patients. Surg Endosc 2007; 21: 994 – 997

15 *Cotton P*, *Williams C*. Practical gastrointestinal endoscopy: the fundamentals. Oxford: John Wiley & Sons; 2008: 88 – 93

16 *Shah SG*, *Thomas-Gibson S*, *Lockett M* et al. Effect of real-time magnetic endoscope imaging on the teaching and acquisition of colonoscopy skills: results from a single trainee. Endoscopy 2003; 35: 421 – 425

17 *Nerup N*, *Preisler L*, *Svendsen MBS* et al. Assessment of colonoscopy by use of magnetic endoscopic imaging: design and validation of an automated tool. Gastrointest Endosc 2015; 81: 548 – 554

18 *Messick SA*. Validity. In: Linn RL, ed. Educational measurement. 3rd edn. New York: Macmillan; 1989: 13 – 103

19 *Downing SM*. Validity: on the meaningful interpretation of assessment data. Med Educ 2003; 37: 830 – 837

20 *Brennan RL*. Generalizability theory. New York: Springer; 2001: 190 – 207

21 *Downing SM*, *Yudkowsky R*. Assessment in health professions education. New York: Routledge; 2009: 119 – 148

22 *Spier BJ*, *Benson M*, *Pfau PR* et al. Colonoscopy training in gastroenterology fellowships: determining competence. Gastrointest Endosc 2010; 71: 319 – 324

23 *Farthing MJ*, *Walt RP*, *Allan RN* et al. A national training programme for gastroenterology and hepatology. Gut 1996; 38: 459 – 470

24 *Downing SM*. Reliability: on the reproducibility of assessment data. Med Educ 2004; 38: 1006 – 1012

25 *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, 1133.e1 – 3

26 *Dancey CP*. Statistics without maths for psychology: using SPSS for Windows. 4th ed. Harlow, UK; New York: Pearson/Prentice Hall; 2007: 168 – 203

27 *Chung YW*, *Han DS*, *Yoo K-S* et al. Patient factors predictive of pain and difficulty during sedation-free colonoscopy: A prospective study in Korea. Dig Liver Dis 2007; 39: 872 – 876

28 *Konge L*, *Larsen KR*, *Clementsen P* et al. Reliable and valid assessment of clinical bronchoscopy performance. Respiration 2012; 83: 53 – 60

29 *Manzone J*, *Tremblay L*, *You-Ten KE* et al. Task- versus ego-oriented feedback delivered as numbers or comments during intubation training. Med Educ 2014; 48: 430 – 440

30 *Korman LY*. Skill and colonoscopy: even the pros take lessons. Gastrointest Endosc 2015; 81: 555 – 556