

Influence of video-based feedback on self-assessment accuracy of endoscopic skills: a randomized controlled trial*



Authors

Michael A. Scaffidi¹, Catharine M. Walsh^{2,3,4}, Rishad Khan¹, Colleen H. Parker¹, Ahmed Al-Mazroui¹, Michael Abunassar¹, Alexander W. Grindal¹, Peter Lin¹, Christopher Wang¹, Robert Bechara⁵, Samir C. Grover¹

Institutions

- 1 Division of Gastroenterology, St. Michael's Hospital, University of Toronto, Toronto, Canada
- 2 Department of Paediatrics, University of Toronto, Toronto, Canada
- 3 Department of Medicine, University of Toronto, Toronto, Canada
- 4 Division of Gastroenterology, Hepatology and Nutrition, Hospital for Sick Children, University of Toronto, Toronto, Canada
- 5 Department of Medicine, Queen's University, Kingston, Ontario, Canada

submitted 26.11.2018

accepted after revision 7.2.2019

Bibliography

DOI <https://doi.org/10.1055/a-0867-9626> |

Endoscopy International Open 2019; 07: E678–E684

© Georg Thieme Verlag KG Stuttgart · New York

eISSN 2196-9736

Corresponding author

Samir C. Grover, MD, MEd, FRCPC, Division of Gastroenterology, St. Michael's Hospital, Department of Medicine, University of Toronto, 16-036 Cardinal Carter Wing, 30 Bond Street, Toronto, ON M5B 1W8, Canada
 Fax: -416-864-5882
samir.grover@utoronto.ca

ABSTRACT

Background and study aims Novice endoscopists are inaccurate in self-assessment of procedures. One means of improving self-assessment accuracy is through video-based

feedback. We aimed to determine the comparative effectiveness of three video-based interventions on novice endoscopists' self-assessment accuracy of endoscopic competence.

Materials and methods Novice endoscopists (performed <20 previous procedures) were recruited. Participants completed a simulated esophagogastroduodenoscopy (EGD) on a virtual reality simulator. They were then randomized to one of three groups: self-video review (SVR), which involved watching a recorded video of their own performance; benchmark review (BVR), which involved watching a video of a simulated EGD completed by an expert; and self- and benchmark video (SBVR), which involved both videos. Participants then completed two additional simulated EGD cases. Self-assessments were conducted immediately after the first procedure, after the video intervention and after the additional two procedures. External assessments were conducted by two experienced endoscopists, who were blinded to participant identity and group assignment through video recordings. External and self-assessments were completed using the global rating scale component of the Gastrointestinal Endoscopy Competency Assessment Tool (GIECAT GRS).

Results Fifty-one participants completed the study. The BVR group had significantly improved self-assessment accuracy in the short-term, compared to the SBVR group ($P = .005$). The SBVR group demonstrated significantly improved self-assessment accuracy over time ($P = .016$). There were no significant effects of group or of time for the SVR group.

Conclusions Video-based interventions, particularly combined use of self- and benchmark video review, can improve accuracy of self-assessment of endoscopic competence among novices.

Introduction

Self-assessment of endoscopic performance by trainees allows for regulation of learning and skill acquisition [1, 2]. The Joint Advisory Group on Gastrointestinal Endoscopy (JAG) recom-

* Meeting presentations: Digestive Disease Week 2017

mends that trainees incorporate self-assessment practices into their self-regulated development [3] and the American Society of Gastrointestinal Endoscopy (ASGE) provides tools for self-assessment of endoscopic performance [4]. To be an effective source of feedback, self-assessment must be accurate.

The endoscopic literature, however, has shown that novices have inaccurate self-assessment [5–8]. A recent cross-sectional study on colonoscopy found that novices have less accurate self-assessments compared to more experienced endoscopists [6], consistent with studies in other procedure-related domains [9, 10]. In addition, a study of simulated polypectomy found a weak correlation between self- and externally-assessed performance scores among novices [5].

Video-based feedback has been proposed to remedy deficiencies in self-assessment ability. Several studies found that allowing novices to review videos of their own performances [11], of a benchmark performance (i. e. video of an expert completing the procedure) [8], or both [12], improved self-assessment accuracy. The impact of video-based interventions on endoscopists' self-assessment accuracy, however, is unclear. Moreover, combined use of both self-video review and benchmark video review has not been investigated in a procedural setting. The aim of this study was to ascertain the comparative effectiveness of three different video-based interventions on self-assessment accuracy of endoscopic competence in esophagogastroduodenoscopy (EGD).

Material and methods

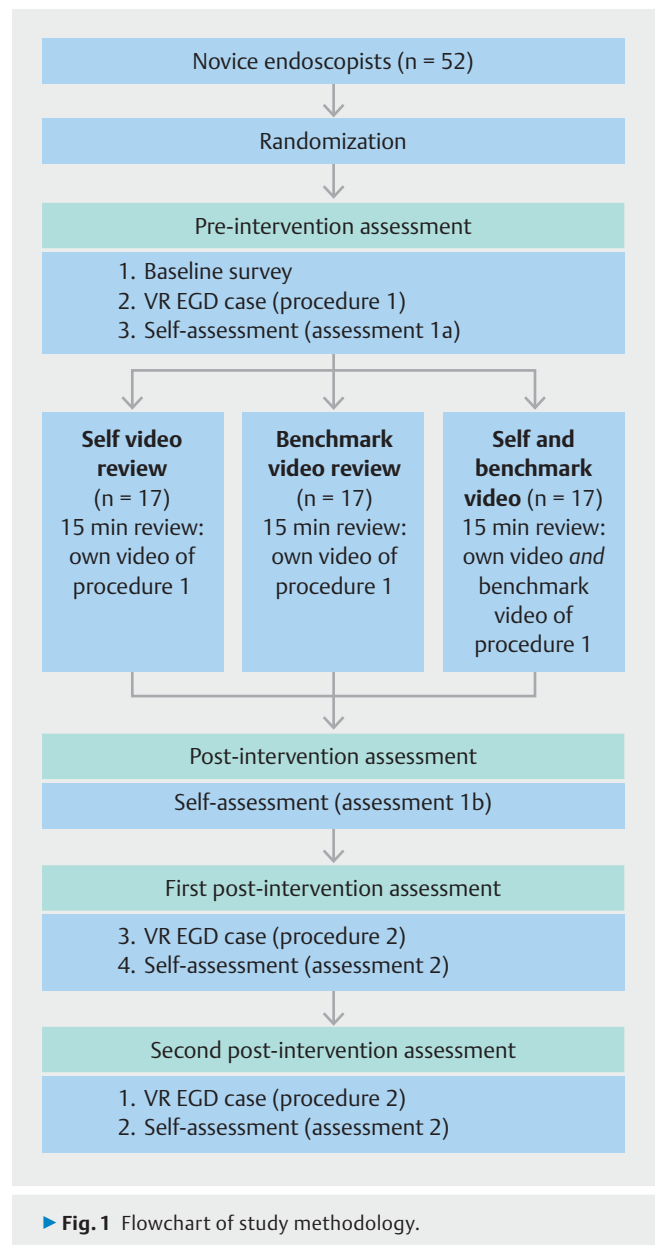
This single-blinded, parallel-arm, prospective randomized controlled trial was conducted at a tertiary care academic center. Approval was granted by the St. Michael's Hospital Research Ethics Board (14–160) and written informed consent was obtained from all participants. Reporting of the findings followed the CONSORT statement [13]. All authors reviewed and approved the final manuscript. No changes to methodology were made after trial commencement.

Participants

One author (MAS) used purposive sampling to recruit novice endoscopists, defined as individuals who had performed fewer than 20 previous EGDs, in the clinical and/or simulated settings [14]. Participants were randomized with an allocation ratio of 1:1:1 to one of the following three groups: (1) self-video review (SVR); (2) benchmark video review (BVR); or (3) self- and benchmark video review (SBVR). Randomization was conducted by one author (RK) using a sealed envelope technique. The random allocation sequence was generated by another author (CW). It was not possible to blind participants to their assigned group.

Procedure

The study methodology is summarized in ► **Fig. 1**. The EndoVR endoscopy simulator was used for all assessments (CAE Healthcare Canada, Montreal, Quebec, Canada). This simulator models an EGD by using an endoscope that is inserted into a computer-based module and displays the esophageal lumen of a



► **Fig. 1** Flowchart of study methodology.

virtual patient on a screen. This simulator was chosen for offering a wide range of EGD cases with variable difficulty and complexity [14, 15]. Two EGD cases were used during testing: Case 1, which represented a 42-year-old male with epigastric pain and a pre-pyloric ulcer; and Case 2, which represented a 41-year-old female with dysphagia and esophageal candidiasis.

Pre-intervention assessment

All participants completed a written questionnaire to collect information on demographic and background characteristics, including age, sex, level of training, and previous experience with endoscopic procedures. Each participant completed an EGD case on the VR simulator (Case 2). A maximum of 15 minutes was allotted for case completion. All participants were video recorded during each of their procedures (as described below).

Participants received no external feedback regarding their performance during the assessments.

Video-based interventions

After completion of the first case, participants received a video-based intervention, according to the group to which they were randomized. The SVR, BVR and SBVR groups were all modeled on the mode of video delivery used in previous studies using self-video review [11], benchmark video review [11] and combined self-review with benchmark review [12].

SVR group

The SVR group was provided with access to footage of their own performance of their first EGD case. Participants had 15 minutes to review the video and could cue forward and backward at their own discretion.

BVR group

The BVR group was provided with access to a benchmark video of the simulated EGD case (Case 2) which featured a demonstration of the task as completed by an experienced endoscopist (>500 endoscopic procedures). Participants had 15 minutes to review the video and could cue forward and backward at their own discretion.

SBVR group

The SBVR group was provided access to footage of their own performance and the benchmark performance during a 15-minute period. They could cue forward and backward and switch between the videos at their own discretion.

Post-intervention assessments

After completion of their assigned video-based intervention, each participant then completed the same simulated EGD case (Case 2) again, followed by a new case (Case 1). A maximum of 15 minutes was allotted for completion of each case. As before, all performances were recorded.

Assessment tools

Performances of the simulated EGD procedures were assessed using the Gastrointestinal Endoscopy Competency Assessment Tool (GiECAT), a direct observational assessment tool, with strong evidence of reliability and validity in the clinical [16, 17] and simulated settings [18, 19]. The GiECAT is composed of a global rating scale (GRS) and structured checklist. Only the GRS component of the GiECAT was used as the items on the GRS are transferable across endoscopic procedures [20]. The GRS assess seven domains (technical skill; strategies for scope advancement; visualization of mucosa; independent procedure completion (need for assistance); knowledge of procedure; interpretation and management of findings; and patient safety) using a 5-point Likert scale with descriptive anchors reflective of the degree of autonomy demonstrated by the endoscopist. Ratings of the seven items on the GRS are totaled to generate scores from 7 to 35. Percentage scores can also be calculated.

Assessments

Video recordings

All three simulated EGD cases performed by each participant were recorded. The protocol for videotaping and editing the video feed of the endoscope's intraluminal view and the endoscopist's hands was adapted from a previous study [21]. Segments of audio and/or video that identified the endoscopist were edited to ensure anonymity. In addition, participants' video review period was video-recorded, which allowed for calculation of the time spent viewing the assigned video(s).

External assessment

Video recordings of all three EGD cases were assessed by two blinded raters (experienced endoscopists who had completed >500 procedures) using the GiECAT GRS. Raters were asked to watch each video in its entirety and to use the full range of responses. A second rater was employed to ensure adequate interrater reliability and was thus assigned a subset of the videos.

Self-assessment

Participants assessed their own performance at four time points: "Assessment 1a," which was immediately after their first simulated EGD; "Assessment 1b," which was immediately after completion of their assigned intervention (SVR, BVR, SBVR) and involved a reappraisal of their first procedure; "Assessment 2," which was immediately after completion of their second simulated EGD; and "Assessment 3," which was immediately after completion of their third simulated EGD. Participants self-assessed their EGD performance at each time point using the GiECAT GRS and were asked to use the full range of responses. The time-period from Assessment 1a to Assessment 3 was no more than 1 hour, as participants were allotted a maximum of 15 minutes to complete each EGD.

Outcome measures

We determined the between- and within-group impacts of the three video-based interventions on self-assessment accuracy for simulated EGD. Self-assessment accuracy was determined by comparing ratings assigned by participants and external assessors on the GiECAT GRS.

Sample size calculation

Based on previous work using educational interventions in endoscopic training, we estimated that 17 participants would be required per group [18]. Under this assumption, we recruited a total of 51 participants.

Statistical analysis

Demographic variables, endoscopic experience, and time spent on the respective video interventions were summarized using descriptive statistics. Calculation of GiECAT GRS percentage score was adapted from the original paper [17]. The mean of the assessments from the two raters were used; when both were not available, only one rater's score was used. The second rater assessed the performance of 31 participants (61%). For these performances, the interrater reliability of the video-

► **Table 1** Endoscopist participant demographic characteristics and previous endoscopic experience.

Characteristic		SVR group (n = 17)	BVR group (n = 17)	SBVR group (n = 17)
Age (years), median (IQR)		27.0 (8.0)	27.0 (8.0)	27.0 (7.0)
Sex	Male, no. (%)	14 (82.4)	12 (70.6)	12 (70.6)
	Female, no. (%)	3 (17.6)	5 (29.4)	5 (29.4)
Level of training or practice	Medical student, no (%)	6 (35.3)	5 (29.4)	6 (35.3)
	Resident, no. (%)	8 (47.1)	11 (64.7)	9 (52.9)
	Staff/attending, no. (%)	3 (17.6)	1 (5.9)	2 (11.8)
Hand dominance	Right, no. (%)	17 (100)	15 (88.2)	17 (100)
	Left, no. (%)	0 (0)	2 (11.8)	0 (0)
Endoscopic experience	Number of previous colonoscopies completed, median (IQR)	0 (2.0)	0 (2.0)	0 (0)
	Number of previous EGDs completed, median (IQR)	0 (4.0)	0 (2.0)	0 (0)

BVR, benchmark video review; EGD, esophagoduodenoscopy; IQR, interquartile range; SVR, self-video review; SBVR, self- and benchmark video review

based expert assessments was calculated using the intraclass correlation coefficient ($ICC_{2,1}$), 2-way random-effect model for average measures.

To determine self-assessment accuracy, two approaches were used based on recommendations from the method comparison literature [22] and from a previous study examining self-assessment accuracy of endoscopic competence [6]. First, to determine overall self-assessment accuracy of participants at baseline (i.e. prior to the intervention), the $ICC_{1,1}$ (1-way random-effects model for both single measures [individual rater] and average measures [average of 2 raters' scores]) was calculated using the GiECAT GRS scores assigned by external assessors and by participants for a single EGD procedure. Second, a Bland-Altman analysis was used to compare agreement between self- and externally-assessed GiECAT GRS scores at baseline (i.e. assessment 1a) among the three groups [23].

To evaluate the impact of the video-based interventions on self-assessment accuracy, absolute difference scores (ADS) between externally- and self-assessed GiECAT GRS scores among the three groups were determined. To determine if there was a between-group effect, Kruskal-Wallis tests were completed for the ADS among the three groups at each assessment (Assessment 1a, 1b, 2, 3). To determine if there was a within-group effect, Friedman tests were completed for the ADS over the four assessment time points (Assessment 1a, 1b, 2, 3) for each group.

All analyses were conducted using SPSS 20 (IBM, Armonk NY). Interpretation of the ICC followed suggested guidelines, wherein values 0.21–0.40 are considered “fair,” 0.41–0.60 “moderate,” 0.61–0.80 “substantial,” and >0.80 “almost perfect” [24]. Any significant effects on the Kruskal-Wallis and Friedman tests were further analyzed using Mann-Whitney U tests and Wilcoxon signed-rank tests, respectively. Multiple post hoc comparisons were corrected for using the Dunn-Sidak adjustment, following a pairwise approach [25]. Effect size was

calculated using eta squared (η^2) for Kruskal-Wallis tests and Kendall's W for Friedman tests [26]. For all statistical tests, an alpha of 0.05 was set as the cut-off for statistical significance.

Results

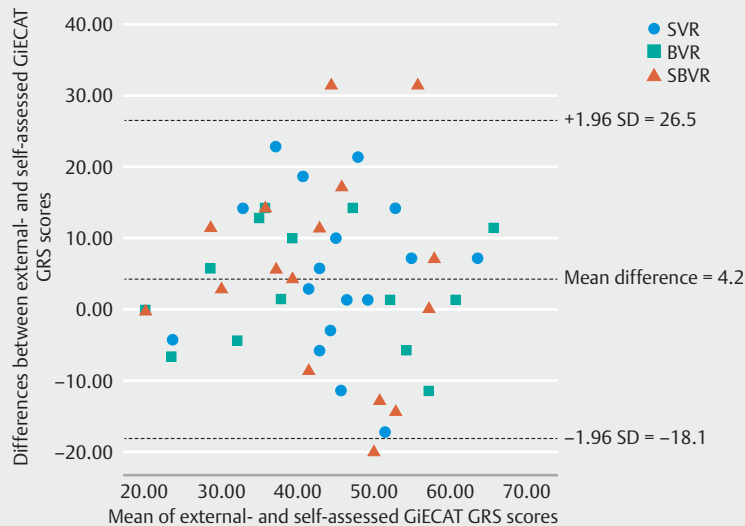
A total of 51 novice endoscopists were randomized and completed the study. Participant demographics and endoscopic experience are outlined in ► **Table 1**. Inter-rater reliability for the two video-based external reviewers was good, as indicated by an $ICC_{2,1}$ value of 0.73 (95% CI, 0.43–0.87), 0.88 (0.74–0.94), and 0.73 (0.42–0.87) for assessments 1a, 2, and 3, respectively.

Median time spent on the self-video review and on the benchmark video was 14 minutes, 34 seconds (IQR: 4 min, 3 s) and 13 minutes, 12 seconds (IQR: 6 min, 44 s) by the SVR and BVR groups, respectively. Median time spent by the SBVR group on the self-video review and on the benchmark video was 8 minutes, 1 second (IQR: 4 mins, 48 s) and 6 minutes, 47 seconds (IQR: 2 mins, 6 s), respectively.

Self-assessment accuracy

Baseline

Overall, there was moderate agreement between the external and self-assessments for the GiECAT GRS at baseline (i.e. assessment 1a), as evidenced by an $ICC_{1,1}$ (average measure) of 0.74 (95% CI, 0.48–0.88). In the Bland-Altman analysis, the mean of the differences between externally assessed and self-assessed GiECAT GRS scores was 4.2 ($SD=11.4$) (► **Fig. 2**). All but three data points fell within the 95% limits of agreement, as two participants in the SBVR group fell above the upper limit and one participant in the SBVR group fell below the lower limit. There were no systematic differences between the three groups.



► Fig. 2 Bland-Altman plot.

► **Table 2** Absolute difference scores between external- and self-assessed GiECAT GRS scores for participants in the SVR, BVR and SBVR groups. Values are median ratings with the interquartile range in parentheses.

Procedure ¹	Absolute difference percentage Score (%)			P value ²		
	SVR	BVR	SBVR	SVR-BVR	SVR-SBVR	BVR-SBVR
1a	7.1 (12.1)	5.7 (10.0)	11.4 (9.6)	NS	NS	NS
1b	10.0 (13.6)	5.7 (7.9)	14.3 (14.3)	NS	NS	0.005 ³
2	5.7 (13.6)	7.1 (13.2)	10.0 (12.5)	NS	NS	NS
3	14.3 (14.3)	14.3 (12.5)	6.4 (18.2)	NS	NS	NS

GiECAT, Gastrointestinal Endoscopy Competency Assessment Tool; GRS, global rating scale; NS, Not Significant (at $P < .05$)

¹ Note that procedures 1a and 1b correspond to the periods before and after completing the assigned video-based intervention, respectively.

² Significant differences between groups ($P < .05$). Post-hoc comparisons were carried out by using Mann Whitney U tests.

³ Denotes a significant difference ($P < .05$)

Effects of video-based interventions

The ADS for all assessments using the GiECAT GRS among the three groups is presented in ► **Table 2**. There was a significant effect of group for the absolute difference of externally- and self-assessed GiECAT GRS scores for procedure 1b (Kruskal-Wallis chi-squared = 9.78₂, $P = .008$, $\eta^2 = 0.17$). There were no significant differences for procedure 1a (Kruskal-Wallis chi-squared = 4.12₂, $P = .127$), procedure 2 (Kruskal-Wallis chi-squared = 1.60₂, $P = .449$), or procedure 3 (Kruskal-Wallis chi-squared = 1.13₂, $P = .519$). Post hoc analysis indicated that the BVR group had a significantly smaller ADS compared to the SBVR group on procedure 1b ($P = .005$). There were no other significant differences.

There was a significant effect of time for the BVR group (Friedman chi-squared = 9.40₂, $P = .024$, $\eta^2 = 0.06$) and for the SBVR group (Friedman chi-squared = 10.35₂, $P = .016$, $\eta^2 = 0.07$). There was no significant effect of time for the SVR group

(Friedman chi-squared = 1.43₂, $P = .698$). Post hoc analysis indicated that the BVR group had a significantly higher ADS on assessment 3 compared to assessment 1b ($P = .030$) and the SBVR group had a significantly lower ADS on assessment 3 compared to assessment 1b ($P = .016$). There were no other significant differences.

Discussion

We report the first study to assess the comparative effectiveness of various video-based interventions aimed at improving self-assessment accuracy of procedural skills. We found that benchmark video review on its own was beneficial in the short term only, while self-video review in isolation was not. In addition, we found that benchmark video review paired with self-video review improved self-assessment accuracy over time. Self-assessment is an essential skill wherein individuals monitor their learning and performance [27]. Accurate self-assessment

involves adequate agreement between one's own assessment when compared to an external standard [28]. Although novice endoscopists have been shown to have inaccurate self-assessment [29], our findings suggest that their abilities can be enhanced using video-based interventions.

There are several potential explanations for our results. The SVR group may have self-assessed themselves based on an overall impression of their performance, which did not change with self-video review alone as they had no appropriate external standard against which they could compare their own performance [30]. The benchmark video, on the other hand, likely provided an advantage to the BVR group as novices could use the expert performance to help identify flaws in their own endoscopic skills. This is consistent with a previous study in sigmoidoscopy, in which general surgery residents had improved self-assessment accuracy after watching an expert performance [8].

Given our finding that the BVR group had improved self-assessment accuracy compared to the SBVR group, in the short term, we hypothesize that the benefit of the benchmark video alone may be attributable to the lower cognitive load required to process one video. Conversely, participants in the SBVR group may have initially been challenged to effectively process both videos within the allotted time. With time, however, participants in the SBVR group may have been better able to reflect on their own video and the degree to which their performance met the benchmark standard, thereby informing their self-assessment. The finding that self-video review is only beneficial when combined with benchmark video-review is commensurate with previous work on this subject [31]. In addition, video-based feedback appears to mitigate the Dunning-Kruger effect. According to this effect, novices are unaware of their own skill deficiencies wherein the least competent are more likely to overestimate their level of performance [32]. Accurate self-assessment requires appropriate external standards for measuring one's performance and the ability to judge the extent to which one's own performance meets those standards. Providing novices with a video of their own performance as well as a benchmark performance likely enhances self-assessment accuracy as it provides trainees with high-quality data which they can use to interpret their own performance and compare it to an explicit standard.

This study has several limitations. First, we used the GiECAT GRS to evaluate EGD performance, as there are no EGD-specific assessment tools with strong validity evidence. Although the GiECAT GRS has been validated for use in colonoscopy, it is lacking comprehensive evidence of validity for EGD [17]. In addition, we did not use a control group (i. e. no video intervention), so we are unable to determine if participants' self-assessment accuracy would have improved over time with no intervention. A previous study, however, suggested that a control group would show no improvement [8]. Finally, our study evaluated the self-assessment accuracy of participants within a single day. It is possible that differences between groups would change over a longer observational period.

Overall, video-based interventions can improve accuracy of self-assessment of endoscopic competence among novices. In

particular, benchmark performances in combination with a self-video review, may help to better inform these assessments. There are several implications of our findings. First, video-based interventions may be integrated into existing endoscopic training curricula [18, 19, 33] to facilitate recognition of performance deficits among novices. Video-recording has demonstrated benefits as a tool for external assessment and debriefing [34], and, based on our findings, it may also be used to improve learning by promoting accurate self-assessment. Ensuring trainees have accurate perceptions of their endoscopic competence may facilitate their learning as several studies in the educational literature have demonstrated that trainees are more receptive to feedback and more likely to uptake external feedback if it aligns with their self-perception [35, 36].

Conclusion

Research has shown that it is critical for trainees to have an accurate perception of their abilities as their own opinions, as opposed to external assessments, predominately influence the generation of learning goals [36]. An online compendium of benchmark videos for major endoscopic procedures that feature a variety of presentations and techniques would be a useful resource for novices. The American Society for Gastrointestinal Endoscopy's extensive database of videos could be updated to include annotations of key aspects of the performance in reference to an assessment standard to facilitate self-assessment. Future studies are required to investigate video-based interventions targeting other endoscopic procedures and evaluate their impact on self-assessment accuracy over a longer time period.

Acknowledgements

The authors thank Samantha Scaffidi, Grace Kim, Cindy Tsui, Soha Iqbal, and Shahin Akbari-Nasab for assistance in data acquisition and Roger Chow for administrative assistance.

Competing interests

Dr. Khan received research funding from Abbvie and Ferring Pharmaceuticals. Dr. Parker received educational support from Allergan Canada. Dr. Bechara is a consultant for Olympus. Dr. Grover received grants from Abbvie and Janssen, personal fees from Abbvie, Takeda, and Ferring Pharmaceuticals and is owner and has shares in Volō Healthcare.

References

- [1] Ekkelenkamp VE, Koch AD, Haringsma J et al. Quality evaluation through self-assessment: a novel method to gain insight into ERCP performance. *Frontline Gastroenterol* 2014; 5: 10–16
- [2] 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

- [3] Joint Advisory Group on Gastrointestinal Endoscopy. Joint Advisory Group on Gastrointestinal Endoscopy (JAG) accreditation standards for endoscopy services. 2014: Available at <https://www.thejag.org.uk/downloads/Accreditation/JAG%20accreditation%20standards%20for%20endoscopy%20services.pdf> [Accessed February 26, 2019]
- [4] American Society of Gastrointestinal Endoscopy. Gastrointestinal Endoscopy Self-Assessment Program VIII (GESAP VIII).
- [5] Ansell J, Hurley JJ, Horwood J et al. Can endoscopists accurately self-assess performance during simulated colonoscopic polypectomy? A prospective, cross-sectional study *Am J Surg* 2014; 207: 32–38
- [6] Scaffidi MA, Grover SC, Carnahan H et al. Impact of experience on self-assessment accuracy of clinical colonoscopy competence. *Gastrointest Endosc* 2018; 87: 827–836
- [7] Moritz V, Holme O, Leblanc M et al. An explorative study from the Norwegian Quality Register Gastronet comparing self-estimated versus registered quality in colonoscopy performance. *Endosc Int Open* 2016; 4: E326–E332
- [8] Vyasa P, Willis RE, Dunkin BJ et al. Are general surgery residents accurate assessors of their own flexible endoscopy skills? *J Surg Educ* 2016; 74: 23–29
- [9] Arora S, Miskovic D, Hull L et al. Self vs expert assessment of technical and non-technical skills in high fidelity simulation. *Am J Surg* 2011; 202: 500–506
- [10] Brewster LP, Risucci DA, Joehl RJ et al. Comparison of resident self-assessments with trained faculty and standardized patient assessments of clinical and technical skills in a structured educational module. *Am J Surg* 2008; 195: 1–4
- [11] Ward M, MacRae H, Schlachta C et al. Resident self-assessment of operative performance. *Am J Surg* 2003; 185: 521–524
- [12] Hawkins SC, Osborne A, Schofield SJ et al. Improving the accuracy of self-assessment of practical clinical skills using video feedback The importance of including benchmarks. *Med Teach* 2012; 34: 279–284
- [13] Cheng A, Kessler D, Mackinnon R et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Adv Simul* 2016; 1: 25
- [14] Ekkelenkamp VE, Koch AD, De Man RA et al. Training and competence assessment in GI endoscopy: a systematic review. *Gut* 2016; 65: 607–615
- [15] Triantafyllou K, Lazaridis LD, Dimitriadis GD. Virtual reality simulators for gastrointestinal endoscopy training. *World J Gastrointest Endosc* 2014; 6: 6–12
- [16] Walsh CM, Ling SC, Walters TD et al. Development of the Gastrointestinal Endoscopy Competency Assessment Tool for Pediatric Colonoscopy (GiECATKIDS). *J Pediatr Gastroenterol Nutr* 2014; 59: 480–486
- [17] Walsh CM, Ling SC, Khanna N et al. Gastrointestinal endoscopy competency assessment tool: Reliability and validity evidence. *Gastrointest Endosc* 2015; 81: 1417–1424.e2
- [18] Grover SC, Garg A, Scaffidi MA et al. Impact of a simulation training curriculum on technical and nontechnical skills in colonoscopy: a randomized trial. *Gastrointest Endosc* 2015; 82: 1072–1079
- [19] Grover SC, Scaffidi MA, Khan R et al. Progressive learning in endoscopy simulation training improves clinical performance: a blinded randomized trial. *Gastrointest Endosc* 2017; 86: 881–889
- [20] Ilgen JS, Ma IWY, Hatala R et al. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015; 49: 161–173
- [21] Scaffidi MA, Grover SC, Carnahan H et al. A prospective comparison of live and video-based assessments of colonoscopy performance. *Gastrointest Endosc* 2016; 83: AB116
- [22] Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. *Theriogenology* 2010; 73: 1167–1179
- [23] Ludbrook J. Confidence in Altman-Bland plots: A critical review of the method of differences. *Clin Exp Pharmacol Physiol* 2010; 37: 143–149
- [24] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174
- [25] Dunn OJ. Multiple comparisons using rank sums *American. Technometrics* 1964; 6: 241–252
- [26] Tomczak M, Tomczak E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci* 2014; 1: 19–25
- [27] Falchikov N, Boud D. Student self-assessment in higher education: a meta-analysis. *Rev Ed Res* 1989: doi:10.3102/00346543059004395
- [28] Sargeant J, Armson H, Chesluk B et al. The processes and dimensions of informed self-assessment: a conceptual model. *Acad Med* 2010; 85: 1212–1220
- [29] Scaffidi MA, Grover SC, Carnahan H et al. Impact of experience on self-assessment accuracy of clinical colonoscopy competence. *Gastrointest Endosc* 2018; 87: 827–836
- [30] Vnuk A, Owen H, Plummer J. Assessing proficiency in adult basic life support: student and expert assessment and the impact of video recording. *Med Teach* 2006; 28: 429–434
- [31] Hawkins SC, Osborne A, Schofield SJ et al. Improving the accuracy of self-assessment of practical clinical skills using video feedback—the importance of including benchmarks. *Med Teach* 2012; 34: 279–284
- [32] Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol* 1999; 77: 1121–1134
- [33] Khan R, Scaffidi MA, Walsh CM et al. Simulation-based training of non-technical skills in colonoscopy: protocol for a randomized controlled trial. *JMIR Res Protoc* 2017; 6: e153
- [34] Scaffidi MA, Grover SC, Carnahan H et al. A prospective comparison of live and video-based assessments of colonoscopy performance. *Gastrointest Endosc* 2017; 87: 688–694.e2
- [35] Eva KW, Armson H, Holmboe E et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ Theory Pract* 2012: 171 5–26
- [36] Eva KW, Munoz J, Hanson MD et al. Which factors, personal or external, most influence students' generation of learning goals? *Acad Med* 2010; 85: S102–S105