Title

Effect of an online educational module incorporating real-time feedback on accuracy of polyp sizing in trainees: a randomized controlled trial.

Journal

Endoscopy, 56(6)

Authors

Mun, Elijah
Yen, Timothy
Hochheimer, Camille
et al.

Publication Date

2024-06-01

DOI

10.1055/a-2245-6526

Peer reviewed

HHS Public Access

Author manuscript

Endoscopy. Author manuscript; available in PMC 2024 June 01.

Published in final edited form as:

Endoscopy. 2024 June ; 56(6): 421–430. doi:10.1055/a-2245-6526.

Effect of an online educational module incorporating real-time feedback on accuracy of polyp sizing in trainees: a randomized controlled trial

Elijah J. Mun[1], **Timothy Yen**[1], **Camille J. Hochheimer**[2], **Wyatt Tarter**[2], **Tonya Kaltenbach**[3,4], **Rajesh N. Keswani**[5], **Sachin Wani**[1], **Swati G. Patel**[1,6]

[1]Division of Gastroenterology, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, United States

[2]Department of Biostatistics and Informatics, Center for Innovative Design and Analysis, Colorado School of Public Health, Aurora, United States

[3]Division of Gastroenterology, Department of Medicine, University of California San Francisco, San Francisco, United States

[4]San Francisco Veterans Affairs Medical Center, San Francisco, United States

[5]Division of Gastroenterology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, United States

[6]Department of Medicine, Rocky Mountain Regional Veterans Affairs Medical Center, Aurora, United States

Abstract

Background—Although polyp size dictates surveillance intervals, endoscopists often estimate polyp size inaccurately. We hypothesized that an intervention providing didactic instruction and real-time feedback could significantly improve polyp size classification.

Methods—We conducted a multicenter randomized controlled trial to evaluate the impact of different components of an online educational module on polyp sizing. Participants were randomized to control (no video, no feedback), video only, feedback only, or video + feedback. The primary outcome was accuracy of polyp size classification into clinically relevant categories (diminutive [1–5mm], small [6–9mm], large [≥ 10mm]). Secondary outcomes included accuracy

Corresponding author: Elijah J. Mun, MD, Division of Gastroenterology, Department of Medicine, University of Colorado Anschutz Medical Campus, 1635 Aurora Court, F735, Aurora, CO 80045, United States, elijah.mun@cuanschutz.edu.

Conflict of Interest

T. Kaltenbach is a consultant for Verily Life Sciences and has received research support from Olympus America. R. Keswani is a consultant for Boston Scientific, Neptune Medical, and Medtronic. S. Wani is a consultant for Exact Sciences and Castle Biosciences, and has received research support from Lucid, Ambu, and CDx Diagnostics. S.G. Patel has received research support from Olympus America. E.J. Mun, T. Yen, C.J. Hochheimer, and W. Tarter declare that they have no conflict of interest.

Clinical Trial

Trial Registration: ClinicalTrials.gov | Registration number (trial ID): NCT05846295 | Type of study: Multicenter Randomized Controlled Trial

Supplementary Material

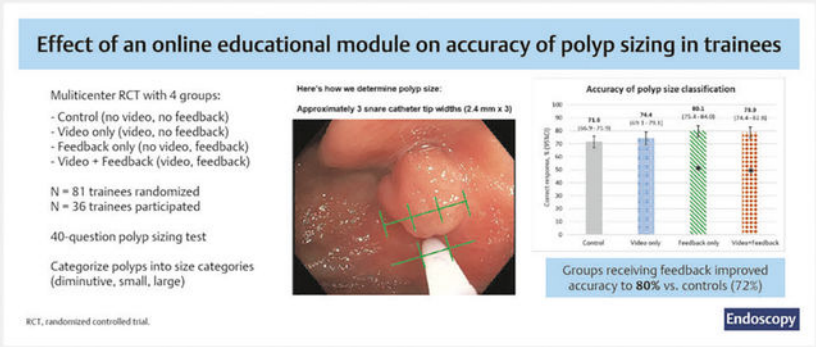
Supplementary Material is available under https://doi.org/10.1055/a-2245-6526

of exact polyp size (inmm), learning curves, and directionality of inaccuracy (over- vs. underestimation).

Results—36 trainees from five training programs provided 1360 polyp size assessments. The feedback only (80.1%, $P = 0.01$) and video + feedback (78.9%, P = 0.02) groups had higher accuracy of polyp size classification compared with controls (71.6%). There was no significant difference in accuracy between the video only group (74.4%) and controls ($P = 0.42$). Groups receiving feedback had higher accuracy of exact polyp size (inmm) and higher peak learning curves. Polyps were more likely to be overestimated than underestimated, and 29.3% of size inaccuracies impacted recommended surveillance intervals.

Conclusions—Our online educational module significantly improved polyp size classification. Real-time feedback appeared to be a critical component in improving accuracy. This scalable and no-cost educational module could significantly decrease under- and overutilization of colonoscopy, improving patient outcomes while increasing colonoscopy access.

Graphical Abstract



Introduction

Although colorectal cancer (CRC) is the third most common cancer and second leading cause of cancer mortality in the United States [1], it can be prevented by removing precancerous polyps [2]. The number, size, and histology of polyps removed are associated with future CRC risk; thus, these factors inform the US Multi-Society Task Force on Colorectal Cancer guidelines on polyp surveillance [3]. While the number and histology of polyps are objective assessments, polyp sizing is subjective, determined by an endoscopist's visual estimation. Notably, endoscopists often estimate polyp size inaccurately [4] and exhibit measurement bias whereby they are more likely to report certain polyp sizes (i.e. 2 mm) over others (i.e. 9 mm) [5]. Improper classification of polyp size has been shown to lead to inappropriate surveillance interval recommendations, potentially in up to 35.2% of cases of polyp mis-sizing [6], thus exposing patients to undue procedures and their potential harms. Various tools have attempted to standardize polyp sizing [7, 8, 9, 10], but these have had modest results and can be costly, time consuming, and cumbersome.

The snare is a tool ubiquitously found across all centers that perform endoscopy, is used in the vast majority of polypectomies, and has prespecified size dimensions that are provided

on individual packaging labels [11]. Thus, it has the potential to be leveraged during colonoscopy to provide accurate polyp size estimations.

There is a need for open-access educational tools to train endoscopists to accurately estimate polyp size into clinically relevant categories, including diminutive (1–5mm), small (6–9mm), and large (≥10mm) polyps, because these categories inform polypectomy method [11] and surveillance intervals [3]. Accurate polyp sizing is especially important for minimizing colonoscopy overutilization in the context of already limited colonoscopy capacity [12] and recent expansion of CRC screening to younger individuals [13].

Prior work has shown that structured and real-time feedback improves long-term knowledge retention [14]. Specifically, feedback is vital in acquiring endoscopic skills, such as optical diagnosis of colorectal polyps [15], competency in polypectomy technique [16], recognition of Barrett's esophagus-related neoplasia [17], and lesion detection in capsule endoscopy [18].

Based on this, we hypothesized that polyp size estimation by gastroenterology trainees could be improved using a video didactic incorporating case-based instruction coupled with iterative, real-time feedback. To test this hypothesis, we developed an online educational module to teach trainees how to estimate polyp size by using the prespecified dimensions of a snare. Our primary aim was to determine the impact of the different components of our module (video instruction, feedback) on accuracy of polyp size classification by trainees into clinically relevant categories (diminutive 1–5mm; small 6–9mm; large ≥10mm).

Methods

Study design

This prospective, multicenter, randomized controlled trial was completed over a 2-week period in November 2021.The study was approved by the Colorado Multiple Institutional Review Board. We used the CONSORT checklist when writing our report (see Table 1s in the online-only Supplementary material) [19].

Study participants

A total of 81 trainees (fellowship years 1–4) from five Accreditation Council for Graduate Medical Education-accredited gastroenterology fellowship training programs from diverse regions and training settings in the USA were sent email invitations to participate. Participants provided informed consent and self-reported baseline demographic data.

Study materials (ESTIMATE module)

The Estimating Polyp Size with Snare Tool to Improve Measurement Accuracy for Trainee Education (ESTIMATE) online educational module was developed by the study team between September 2020 and August 2021. Based on the established principles of interactive, case-based didactic instruction in a "flipped" classroom format in medical education [20] and the benefit of real-time feedback on acquisition of endoscopic skills [15, 17, 18], we designed ESTIMATE to be a user-friendly, interactive, web-based module consisting of two educational components: 1) video instruction and 2) real-time feedback

during a polyp size assessment test. Video instruction was provided in a 5-minute didactic clip (available at https://youtu.be/E4tGL3LLV0Q) narrating the importance of accurate polyp size classification and demonstrating proper technique for positioning the snare for polyp size estimation. Several in vivo case examples, with opportunity for trainee interaction and response, were included.

The polyp size assessment test included 40 polyp sizing questions (see Appendix 1s for examples). Based on prior studies assessing the impact of various sizing interventions (mean of 11 polyps, range of 4–22 polyps) [4, 7, 8, 10], and on the well-documented effect of exam length on cognitive fatigue and performance [21], we included 40 polyp questions in our test to ensure there were enough questions to reliably assess performance, without risking test fatigue. To mirror polyp size prevalence observed in routine colonoscopy [22], 65.0% of the polyps were diminutive, 22.5% were small, and 12.5% were large. Each question contained a still image of a polyp with an adjacent snare and asked participants to classify the polyp as diminutive (1–5mm), small (6–9mm), or large (≥10mm), as well as to give the exact polyp size inmm. Participants were also asked to provide their confidence level (high or low) with each response in order to highlight the potential impact of size responses on recommended surveillance intervals. High confidence was defined as a sufficient level of certainty in polyp size to commit to a surveillance interval recommendation based on the response. Low confidence was defined as an insufficient level of certainty in polyp size to commit to a surveillance interval recommendation based on the response.

Real-time feedback on correct polyp size was provided immediately after a submitted response in the form of a series of images with overlying graphics depicting polyp size relative to the adjacent snare, with a final image of the resected polyp affixed to a cork board with an adjacent ruler (Appendix 2s).

When possible, a large margin of normal tissue was intentionally taken during polypectomy to allow normal tissue to be distinguished from polyp tissue for accurate measurement. All tissue needed to be whole and unaltered in order to be considered. Specimens that had become fragmented while being suctioned through the scope channel were not used. We used the immediate post-resection, pre-formalin fixation polyp to determine the correct size. There was 100% concordance between in vivo measurements for polyp size categories (diminutive, small, large) using the snare and ex vivo cork board measurements, supporting validity of this method. Feedback images were provided in continuous fashion after each response as participants progressed through the test.

Randomization

We randomized participants using the Research Electronic Data Capture (REDCap) randomization feature to one of four study groups: control, video only, feedback only, or video + feedback (Fig. 1). We chose a randomized controlled trial design over a pre/post analysis, which is commonly used in educational trials, as we anticipated some degree of learning with repetition while progressing through the 40-question assessment test [23, 24], and thus may have encountered difficulty distinguishing between post-test improvements as a consequence of repetition versus the intervention itself. To understand the impact of the key educational components of our module (video instruction, feedback), we opted

for participant-level randomization in 1:1:1:1 fashion to one of the four groups. The control group completed the 40-item polyp size assessment test without receiving any video instruction or feedback. The video only group received video instruction and completed the assessment, but did not receive any feedback. The feedback only group received feedback after each question, but did not watch the video. Finally, the video + feedback group received both video instruction and feedback as they completed the assessment. Participants could only complete the module in one session as responses were not saved.

Data collection and reporting system

All data were collected and managed using REDCap. All participants received an anonymous and unique REDCap link and completed the study on their own personal devices. Links expired after the 2-week time allotment. Nonresponders were sent email reminders to complete the module three times over the 2-week study period.

All authors had access to the study data and reviewed and approved the final manuscript.

Study outcomes

Our primary outcome was accuracy of polyp size classification into clinically relevant categories: diminutive (1–5mm), small (6–9mm), and large (≥10mm). Secondary outcomes were accuracy of exact polyp size inmm, cumulative accuracy of polyp size classification over the 40-question test (to plot learning curves), confidence level (high or low) of polyp size classification, and directionality of inaccuracy (size overestimation vs. underestimation), including the exact distribution of overestimated polyps (proportion of diminutive polyps upsized as small, diminutive polyps upsized as large, and small polyps upsized as large) and underestimated polyps (proportion of large polyps downsized as small, large polyps downsized as diminutive, and small polyps downsized as diminutive), both overall and by randomization group. When plotting learning curves, a lead-in period of five questions was chosen to ensure cumulative accuracy was reflective of true performance over time.

Statistical analysis and sample size considerations

Bivariable analyses were performed to identify potential differences between groups. Binomial regressions were conducted to detect differences between the control and intervention groups in overall accuracy of polyp size by size classification and exact size inmm, with the outcome being the number of correct responses out of total responses for each participant. Generalized linear mixed models were used to assess differences in confidence including a random participant effect to account for multiple responses within each participant. These models included either randomization group or polyp size category as fixed effects. Models were evaluated using a type I error rate of 0.05 using two-sided hypotheses. Analyses were conducted using R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria). In order to achieve 80% power to detect a 20% difference in the number of correct answers between the intervention groups and the control group, which was our expected difference based on prior polyp sizing studies with similar effect size [7, 8, 9, 10], we estimated 12 trainees would need to complete the module within each group (48 total) using analysis of variance.

Results

Baseline characteristics

A total of 81 trainees were invited and 36 agreed to participate (44.1% response rate). Participants were randomized to one of the four study groups: control (n=10), video only (n = 8), feedback only (n = 9), video + feedback (n=9) (Fig. 1). There were no significant differences between the groups in sex, training setting or region, career plans, annual colonoscopy volume, or days of endoscopy performed per week (Table 1). There was balanced representation of trainees across all years of fellowship training, proportional to the number of available fellows per class. Most participants (77.8%) reported they had not received any prior instruction on polyp sizing.

Primary outcome

Overall, 36 participants completed a mean of 37.8 of the 40 polyp questions for a total of 1360 responses overall. Accuracy of polyp size classification was significantly higher for the feedback only (80.1%, $P = 0.01$) and video + feedback (78.9%, $P = 0.02$) groups compared with the control group (71.6%), but no different when the control group was compared with the video only group (74.4%, $P = 0.42$) (Fig. 2a).

Secondary outcomes

Similarly, accuracy of exact polyp size inmm was significantly higher for the feedback only (42.6%, $P < 0.001$) and video + feedback (44.2%, $P < 0.001$) groups compared with the control group (29.2%), but no different when the control group was compared with the video only group (32.7%, $P = 0.33$) (Fig. 2b).

Learning curves depicting cumulative accuracy of polyp size classification over the 40-question polyp size assessment test by group are shown in Fig. 3. Following a lead-in period of five questions, the feedback only and video + feedback groups visually demonstrated higher peak cumulative accuracy of polyp size classification compared with the control group throughout the majority of the module, reaching a peak cumulative accuracy of 81.6% and 78.9%, respectively. The video only and control groups reached a peak cumulative accuracy of 77.5% and 72.4%, respectively.

The confidence level of polyp size classification did not vary significantly between the four groups ($P = 0.79$). The feedback only group reported the lowest raw proportion of high confidence responses (74.5%) out of all four groups despite having the highest accuracy of polyp size classification. The control and video + feedback groups tied for the highest raw proportion of high confidence responses (80.6%), despite the control group having the lowest accuracy of polyp size classification (Table 2). Participants were significantly more confident in estimating diminutive polyps (87.9% adjusted high confidence) compared with estimating small (73.1% adjusted high confidence) or large (78.0% adjusted high confidence) polyps ($P < 0.01$).

Overall, trainees incorrectly estimated 23.8% of polyps (324/1360). Of the incorrect estimations, 73.8% (239/324) were size overestimations and 26.2% (85/324) were size

underestimations (Fig. 4). In total, 29.3% of incorrect estimations would have changed recommended surveillance intervals: 18.2% (59/324) by upsizing a diminutive or small polyp to a large polyp, and 11.1% (36/324) by downsizing a large polyp to a small or diminutive polyp. By group, the control (86.0%) and video only (80.3%) groups had a higher proportion of size overestimations compared with the feedback only (61.5%) and video + feedback (60.5%) groups. The control (28.0%) and video only (18.4%) groups additionally had a higher proportion of size overestimations that would have changed recommended surveillance intervals compared with the feedback only (6.2%) and video + feedback (14.5%) groups.

Discussion

In this multicenter randomized controlled trial, trainees who viewed a didactic video on polyp sizing and received real-time feedback on an assessment test, as well as those who only received feedback on the assessment test, had significantly higher accuracy of polyp size classification into clinically relevant categories compared with those who did not receive feedback. Video instruction alone was insufficient to improve polyp sizing, suggesting that feedback is the most crucial component of our module, either alone or paired with video instruction. This finding affirms our hypothesis and is consistent with prior educational studies that have shown the value of active, iterative learning and real-time feedback in medical education [14], particularly with endoscopic procedures [15, 16, 17, 18].

Learning curves were higher and tended to flatten by the end of the test for the feedback only and video + feedback groups, whereas a true plateau for the control group was less apparent, suggesting that feedback accelerates competence. Peak cumulative accuracy was highest in the groups receiving feedback, with the highest peak accuracy (82%) seen in the feedback only group.This suggests that our module length, with feedback, is sufficient to train participants and improve cumulative accuracy of polyp size classification up to a peak threshold of 82%. Although there is no established threshold of competence in polyp sizing, ≥90% accuracy has been used for other endoscopic skills such as optical diagnosis of polyp histology and cecal intubation rate [15, 25]. Although the peak accuracy in our study did not reach 90%, a goal of ≥90% may be aspirational, given prior data that report polyp sizing accuracy rates as low as 11.1% [9]. Whether trainees can reach a higher peak accuracy threshold with repeat sessions of the educational module or with additional polyp questions requires further study.

We did not find any statistically significant differences in the proportion of high confidence diagnoses between groups, though the control group had the highest raw proportion of high confidence characterizations despite having the lowest accuracy. The lack of correlation between a learner's self-confidence in a certain knowledge domain and competence is well described, perhaps most famously by Dunning and Kruger [26]. The Dunning–Kruger effect explains a general sense of unawareness of one's own lack of competence, and has been demonstrated among physicians, and specifically, among physicians performing procedures [27]. Our finding further highlights the need for a more reliable and objective standard for polyp sizing, rather than relying on an endoscopist's self-assessment. It also highlights

the association between polyp size misclassifications and incorrect surveillance interval recommendations, given that we defined confidence as the certainty in a response to commit to a surveillance interval.

Based on our results, inaccurate polyp sizing has clinically relevant implications in up to 29.3% of misclassifications, where a misclassification changes the recommended surveillance interval. Of all misclassifications, 18.2% (59/324) upsized a nonadvanced polyp to an advanced polyp based on size (≥10mm). An estimated 15 million colonoscopies are performed in the USA annually [12] and 7%–11% of cases have a polyp ≥10mm [28]. If 18.2% of these polyps are misclassified as large polyps, this suggests that 189 000–297 000 patients per year may be inappropriately advised to return for 3-year surveillance (instead of 5–10-year surveillance for nonadvanced adenomas/sessile serrated lesions), and guidelines would recommend their first-degree relatives begin colonoscopy screening at age 40, rather than any screening test at age 45.With limited colonoscopy capacity, correcting oversizing of polyps presents an opportunity to decrease overutilization of colonoscopy, minimize unnecessary harms, and save costs. On the other hand, 11.1% (36/324) of all misclassifications downsized an advanced polyp (≥10mm) to a low risk polyp (<10mm). These underestimations and longer surveillance intervals may put patients at risk for post-colonoscopy CRC. Our educational intervention can minimize these misclassifications. As shown in Fig. 4, the control group had the highest proportion of oversizing and clinically significant oversizing (diminutive or small polyp misclassified as large); thus, this module not only improved overall accuracy, but also potentially minimized clinically relevant misclassifications.

Interestingly, although each intervention decreased the absolute number of inaccurate classifications (Fig. 2), based on results presented in Fig. 4, feedback appeared to even out the type of misclassification to a lower proportion of overcalls compared with control and video instruction only. Although our study was not designed to assess whether our interventions result in a statistically significant difference in the directionality of misclassification, this result warrants future study. As our study and others [29] have shown, endoscopists tend to overestimate rather than underestimate polyp size. Perhaps this is due to a tendency of the untrained to err on the side of caution for a subjective decision that carries high clinical importance, which is then lessened after proper training and instruction, leading to a more even distribution of errors. This phenomenon of novices making more overcall errors, rather than undercall errors, has also been observed in other disciplines [30].

Artificial intelligence (AI) has emerged to aid endoscopists in various endoscopic skills, such as polyp detection [31], optical diagnosis and characterization [32], and depth of cancer invasion [33]. Although it is likely that AI assistance may eventually take over polyp sizing, current evidence for this is limited. Kwak et al. reported the first use of AI for polyp sizing in 2022 and found that their AI technology improves polyp size estimation when compared with visual estimation (control) [34]. However, some authors have expressed concerns about the limitations of the technology, as sizing using its bifurcation-to-bifurcation distance measuring method varies depending on bowel preparation quality, amount of air insufflation, and presence of peristalsis [35]. The difficulty of AI technologies in predicting size likely lies in the inherent challenge of accounting for depth

and size of an in vivo three-dimensional object from a two-dimensional still image on a screen. There must be an objective reference point in the same measurement plane as the object intended for sizing (i.e. polyp) that does not succumb to variables that may interfere with accurate sizing (e.g. bowel preparation, amount of air insufflation, bowel peristalsis). Thus, using the snare to estimate polyp size, as described in ESTIMATE, still holds many advantages in today's landscape and may even inform AI-assisted size detection software in the future. As other tools and innovations emerge into routine clinical practice, our tool can be adapted accordingly.

We acknowledge several limitations to our study. As with all educational interventions, it will be vital to demonstrate whether learning gained as a result of our module is durable over time. The ex vivo nature of our study will require further assessment to determine whether results are generalizable to in vivo polyp sizing during endoscopy. We did however find concordance between in vivo measurements using the snare and ex vivo cork board measurements when creating the feedback image deck. Video-based image tests and feedback would have been more ideal for depicting the dynamic process of polyp sizing compared with two-dimensional still photographs, but we found this logistically challenging to execute. There is no current consensus gold standard for actual polyp sizing, but given the well-documented effects of formalin on tissue shrinkage [36], we decided to use the polyp tissue before formalin fixation adjacent to a ruler as our gold standard. Although AI presents the possibility of standardized polyp size assessment, we believe that training endoscopists on important endoscopic skill sets remains critical, as AI assistance should augment, not replace, skills needed to perform high quality colonoscopy. We acknowledge the potential for selection bias and nonresponse bias inherent to studies using web-based invitations; however, based on Institutional Review Board restrictions, we were unable to collect data on nonresponders in order to compare them with responders. Despite reaching out to over 80 trainees for participation, we did not meet our a priori sample size goal. This is not surprising given the clinical demands of gastroenterology fellowship training and the well-established challenge of getting physicians to respond to web-based invitations [37]. Despite this, a type II error was not committed, and the effect size was larger than anticipated, as we found statistically significant differences in our study outcomes. Had this study had a null result, we would have had concerns that the sample size recruited was not large enough to determine whether there truly were no differences between groups or we were simply underpowered to detect these differences.

In conclusion, ESTIMATE, our online educational polyp sizing module, which provides video didactic instruction and real-time iterative feedback, significantly improved trainee accuracy of polyp size classification into clinically relevant size categories compared with controls. Specifically, feedback appeared to be a crucial component of the module. ESTIMATE also helped trainees to sustain higher cumulative accuracy and peak learning compared with controls. ESTIMATE can help standardize polyp sizing, which will consequently impact important clinical decisions, such as surveillance interval recommendations and age at which first-degree relatives should start screening. Improving polyp sizing has the potential to significantly reduce overutilization of colonoscopy, reduce risks associated with unnecessary colonoscopy, and confer cost savings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We would like to thank the AGA Academy of Educators for their Grant support. We also thank the patients and endoscopy lab nurses and technicians at the study's primary site (Rocky Mountain Veterans Affair Medical Center) who played a significant role in the acquisition of polyp images during colonoscopies that heavily assisted the eventual creation of our module.

Funding Information

NIH/NCATS Colorado CTSA Grant | UL1 TR002535

References

[1]. Siegel RL, Miller KD, Fuchs HE et al. Cancer statistics, 2022. CA Cancer J Clin 2022; 72: 7–33 doi:10.3322/caac.21708 [PubMed: 35020204]

[2]. Wallace MB. New strategies to improve polypectomy during colonoscopy. Gastroenterol Hepatol (N Y) 2017; 13: (Suppl. 03): 1–12

[3]. Gupta S, Lieberman D, Anderson JC et al. Recommendations for follow-up after colonoscopy and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2020; 158: 1131–1153.e5 [PubMed: 32044092]

[4]. Chaptini L, Chaaya A, Depalma F et al. Variation in polyp size estimation among endoscopists and impact on surveillance intervals. Gastrointest Endosc 2014; 80: 652–659 [PubMed: 24679658]

[5]. Plumb AA, Nickerson C, Wooldrage K et al. Terminal digit preference biases polyp size measurements at endoscopy, computed tomographic colonography, and histopathology. Endoscopy 2016; 48: 899–908 doi:10.1055/s-0042-108727 [PubMed: 27441685]

[6]. Eichenseer PJ, Dhanekula R, Jakate S et al. Endoscopic mis-sizing of polyps changes colorectal cancer surveillance recommendations. Dis Colon Rectum 2013; 56: 315–321 doi:10.1097/DCR.0b013e31826dd138 [PubMed: 23392145]

[7]. Abdelrahim M, Saiga H, Maeda N et al. Automated sizing of colorectal polyps using computer vision. Gut 2022; 71: 7–9 doi:10.1136/gutjnl-2021-324510 [PubMed: 34266967]

[8]. Goldstein O, Segol O, Gross SA et al. Novel device for measuring polyp size: an ex vivo animal study. Gut 2018; 67: 1755–1756 doi:10.1136/gutjnl-2017-314829 [PubMed: 28935676]

[9]. Jin HY, Leng Q. Use of disposable graduated biopsy forceps improves accuracy of polyp size measurements during endoscopy. World J Gastroenterol 2015; 21: 623–628 doi:10.3748/wjg.v21.i2.623 [PubMed: 25605986]

[10]. Kaz AM, Anwar A, O'Neill DR et al. Use of a novel polyp "ruler snare" improves estimation of colon polyp size. Gastrointest Endosc 2016; 83: 812–816 doi:10.1016/j.gie.2015.08.082 [PubMed: 26382052]

[11]. Kaltenbach T, Anderson JC, Burke CA et al. Endoscopic removal of colorectal lesions – recommendations by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2020; 158: 1095–1129 doi:10.1053/j.gastro.2019.12.018 [PubMed: 32122632]

[12]. Joseph DA, Meester RG, Zauber AG et al. Colorectal cancer screening: estimated future colonoscopy need and current volume and capacity. Cancer 2016; 122: 2479–2486 doi:10.1002/cncr.30070 [PubMed: 27200481]

[13]. Patel SG, May FP, Anderson JC et al. Updates on age to start and stop colorectal cancer screening: recommendations from the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2022; 162: 285–299 [PubMed: 34794816]

[14]. Davila-Cervantes A, Foulds JL, Gomaa NA et al. Experiences of faculty members giving corrective feedback to medical trainees in a clinical setting. J Contin Educ Health Prof 2021; 41: 24–30 doi:10.1097/CEH.0000000000000322 [PubMed: 33290299]

[15]. Patel SG, Rastogi A, Austin G et al. Gastroenterology trainees can easily learn histologic characterization of diminutive colorectal polyps with narrow band imaging. Clin Gastroenterol Hepatol 2013; 11: 997–1003.e1 [PubMed: 23466714]

[16]. Kaltenbach T, Patel SG, Nguyen-Vu T et al. Varied trainee competence in cold snare polypectomy – results of the COMPLETE randomized controlled trial. Am J Gastroenterol 2023: doi:10.14309/ajg.0000000000002368

[17]. Bergman J, de Groof AJ, Pech O et al. An interactive web-based educational tool improves detection and delineation of Barrett's esophagus-related neoplasia. Gastroenterology 2019; 156: 1299–1308.e3 [PubMed: 30610858]

[18]. Postgate A, Haycock A, Thomas-Gibson S et al. Computer-aided learning in capsule endoscopy leads to improvement in lesion recognition ability. Gastrointest Endosc 2009; 70: 310–316 doi:10.1016/j.gie.2008.11.043 [PubMed: 19386301]

[19]. Schulz KF, Altman DG, Moher D. CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010; 152: 726–732 doi:10.7326/0003-4819-152-11-201006010-00232 [PubMed: 20335313]

[20]. Thistlethwaite JE, Davies D, Ekeocha S et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No.23. Med Teach 2012; 34: e421–444 doi:10.3109/0142159X.2012.680939 [PubMed: 22578051]

[21]. Jensen JL, Berry DA, Kummer TA. Investigating the effects of exam length on performance and cognitive fatigue. PLoS One 2013; 8: e70270 doi:10.1371/journal.pone.0070270 [PubMed: 23950918]

[22]. Lieberman D, Moravec M, Holub J et al. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. Gastroenterology 2008; 135: 1100–1105 doi:10.1053/j.gastro.2008.06.083 [PubMed: 18691580]

[23]. Ausubel DP, Youssef M. The effect of spaced repetition on meaningful retention. J Gen Psychol 1965; 73: 147–150 doi:10.1080/00221309.1965.9711263 [PubMed: 14316956]

[24]. Couture M, Lafond D, Tremblay S. Learning correct responses and errors in the Hebb repetition effect: two faces of the same coin. J Exp Psychol Learn Mem Cogn 2008; 34: 524–532 doi:10.1037/0278-7393.34.3.524 [PubMed: 18444753]

[25]. Keswani RN, Crockett SD, Calderwood AH. AGA clinical practice update on strategies to improve quality of screening and surveillance colonoscopy: expert review. Gastroenterology 2021; 161: 701–711 doi:10.1053/j.gastro.2021.05.041 [PubMed: 34334168]

[26]. 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 doi:10.1037//0022-3514.77.6.1121 [PubMed: 10626367]

[27]. Ahmed O, Walsh TN. Surgical trainee experience with open cholecystectomy and the Dunning–Kruger effect. J Surg Educ 2020; 77: 1076–1081 doi:10.1016/j.jsurg.2020.03.025 [PubMed: 32362558]

[28]. Lieberman DA, Holub J, Eisen G et al. Prevalence of polyps greater than 9mm in a consortium of diverse clinical practice settings in the United States. Clin Gastroenterol Hepatol 2005; 3: 798–805 [PubMed: 16234009]

[29]. Anderson BW, Smyrk TC, Anderson KS et al. Endoscopic overestimation of colorectal polyp size. Gastrointest Endosc 2016; 83: 201–208 doi:10.1016/j.gie.2015.06.058 [PubMed: 26318830]

[30]. Mankad K, Hoey ET, Jones JB et al. Radiology errors: are we learning from our mistakes? Clin Radiol 2009; 64: 988–993 [PubMed: 19748004]

[31]. Wang P, Berzin TM, Glissen Brown JR et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019; 68: 1813–1819 [PubMed: 30814121]

[32]. Jin EH, Lee D, Bae JH et al. Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations. Gastroenterology 2020; 158: 2169–2179.e8 [PubMed: 32119927]

[33]. Luo X, Wang J, Han Z et al. Artificial intelligence-enhanced white-light colonoscopy with attention guidance predicts colorectal cancer invasion depth. Gastrointest Endosc 2021; 94: 627–638.e1 [PubMed: 33852902]

[34]. Kwak MS, Cha JM, Jeon JW et al. Artificial intelligence-based measurement outperforms current methods for colorectal polyp size measurement. Dig Endosc 2022; 34: 1188–1195 doi:10.1111/den.14318 [PubMed: 35385184]

[35]. Utsumi T, Horimatsu T, Sano Y et al. Warning from artificial intelligence against inaccurate polyp size estimation. Dig Endosc 2022; 34: 1196–1197 doi:10.1111/den.14364 [PubMed: 35762027]

[36]. Gopalswamy N, Shenoy VN, Choudhry U et al. Is in vivo measurement of size of polyps during colonoscopy accurate? Gastrointest Endosc 1997; 46: 497–502 [PubMed: 9434215]

[37]. So R, Shinohara K, Aoki T et al. Effect of recruitment methods on response rate in a web-based study for primary care physicians: factorial randomized controlled trial. J Med Internet Res 2018; 20: e28 doi:10.2196/jmir.8561 [PubMed: 29422450]

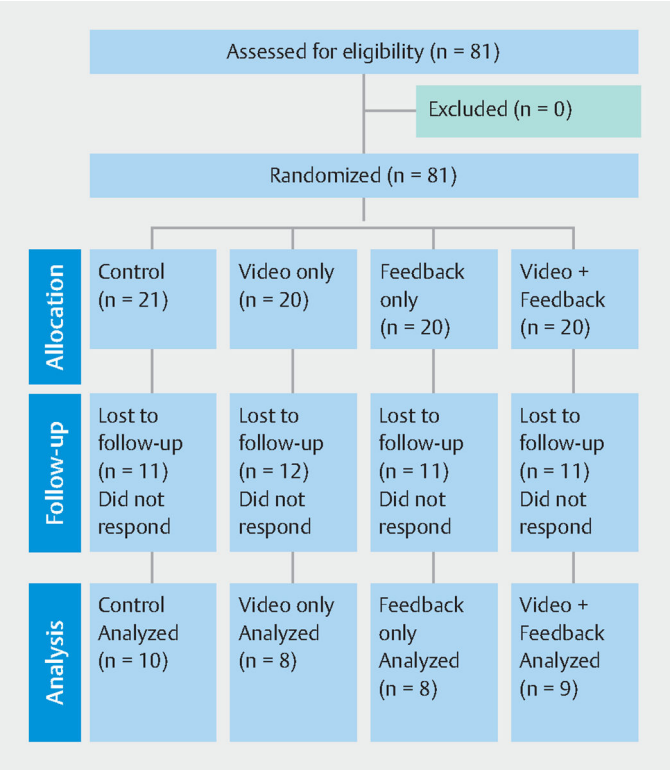


Fig. 1.
Consolidated Standards of Reporting Trials (CONSORT) study schema and randomization.

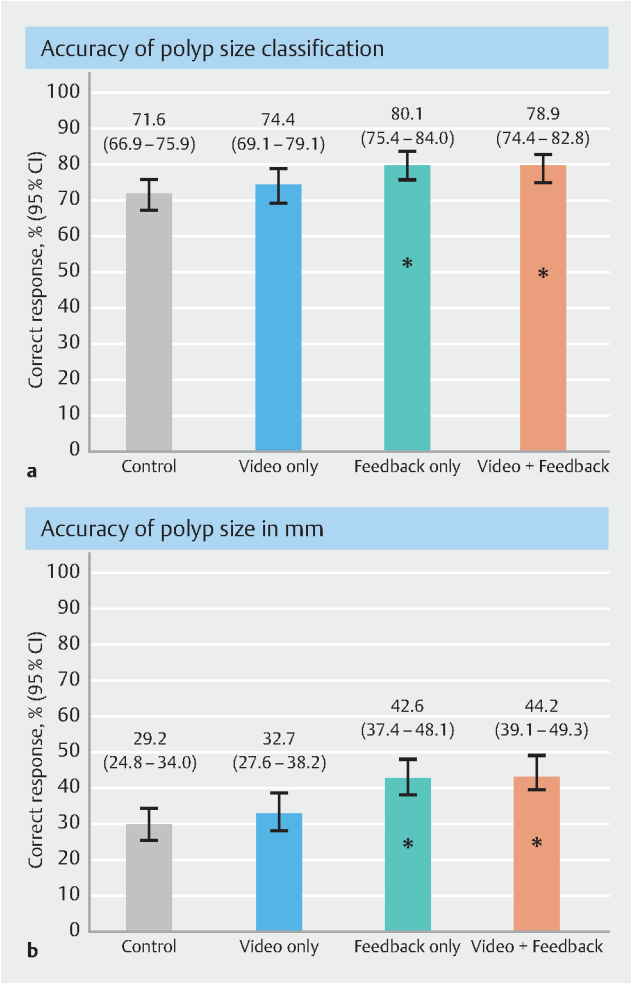


Fig. 2.

Accuracy of polyp size by group. **a** Accuracy by size classification (diminutive 1–5mm; small 5–9mm; large ≥10mm). **b** Accuracy by size inmm. *P <0.05.

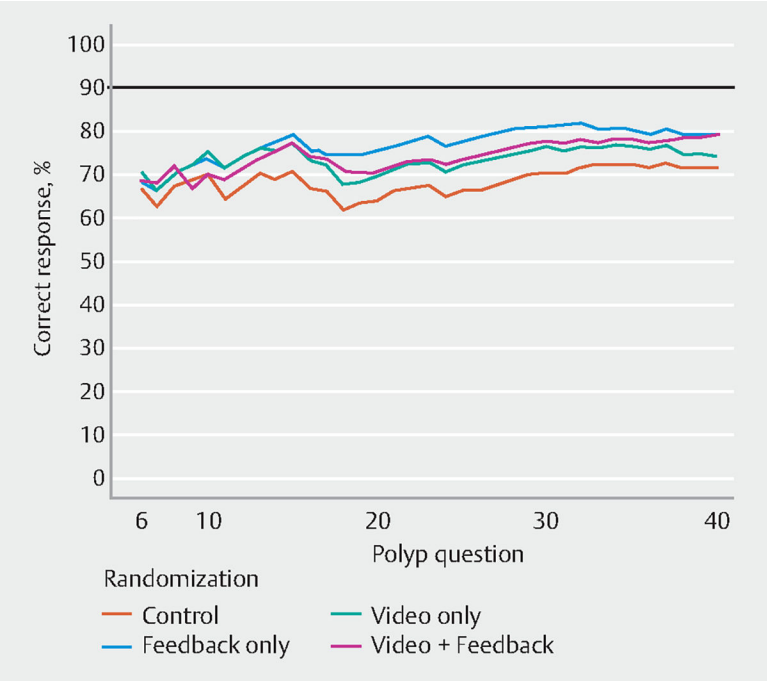


Fig. 3.

Learning curves plotted as cumulative accuracy of polyp size classification over 40-question assessment test by group.

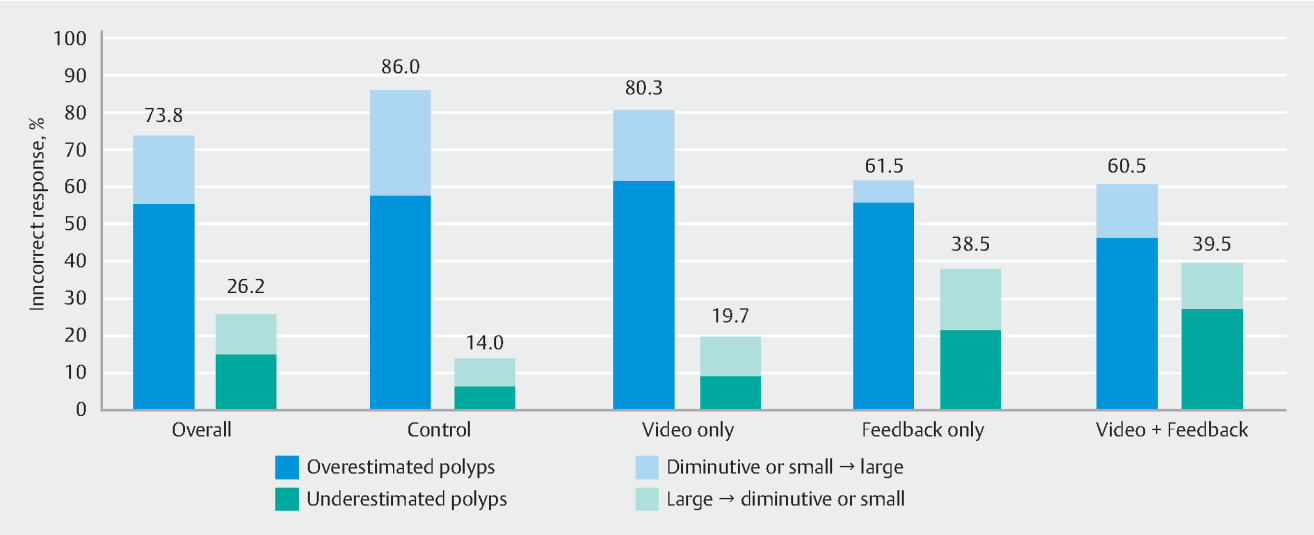


Fig. 4.

Directionality of inaccuracy of polyp size classification (proportion of size overestimations vs. underestimations), overall and by group.

Table 1

Baseline characteristics of study participants.

	Overall	Control	Video only	Feedback only	Video + Feedback	*P* value[1]
Participants, n	36	10	8	9	9	
Sex, n (%)						
Female	14 (38.9)	5 (50.0)	3 (37.5)	3 (33.3)	3 (33.3)	0.86
Male	22 (61.1)	5 (50.0)	5 (62.5)	6 (66.7)	6 (66.7)	
Other/intersex	–	–	–	–	–	
Training setting, n (%)						
University medical center	36 (100)	10 (100)	8 (100)	9 (100)	9 (100)	NA
Safety-net hospital	21 (58.3)	4 (40.0)	5 (62.5)	5 (55.6)	7 (77.8)	0.41
Veterans Affairs hospital	24 (66.7)	6 (60.0)	5 (62.5)	6 (66.7)	7 (77.8)	0.86
Communityhospital	2 (5.6)	1 (10.0)	0 (0.0)	1 (11.1)	0 (0.0)	0.59
Other	–	–	–	–	–	
Training region, n (%)						
West	16 (44.4)	3 (30.0)	5 (62.5)	4 (44.4)	4 (44.4)	0.44
Mountain	14 (38.9)	4 (40.0)	2 (25.0)	4 (44.4)	4 (44.4)	
Midwest	4 (11.1)	3 (30.0)	0 (0.0)	1 (11.1)	0 (0.0)	
South	2 (5.6)	0 (0.0)	1 (12.5)	0 (0.0)	1 (11.1)	
East	–	–	–	–	–	
Training year, n (%)						
Year 1	9 (25.0)	0 (0.0)	2 (25.0)	3 (33.3)	4 (44.4)	0.30
Year 2	13 (36.1)	6 (60.0)	4 (50.0)	1 (11.1)	2 (22.2)	
Year 3	11 (30.6)	3 (30.0)	1 (12.5)	4 (44.4)	3 (33.3)	
Year 4	3 (8.3)	1 (10.0)	1 (12.5)	1 (11.1)	0 (0.0)	
Career plans, n (%)						
Luminal gastroenterology	18 (50.0)	4 (40.0)	5 (62.5)	4 (44.4)	5 (55.6)	0.98
Hepatology	6 (16.7)	2 (20.0)	1 (12.5)	1 (11.1)	2 (22.2)	
Advanced endoscopy	8 (22.2)	3 (30.0)	1 (12.5)	3 (33.3)	1 (11.1)	

	Overall	Control	Video only	Feedback only	Video + Feedback	*P* value[1]
Other/I don't know	4 (11.1)	1 (10.0)	1 (12.5)	1 (11.1)	1 (11.1)	
Intended practice setting, n (%)						
Academic	18 (50.0)	5 (50.0)	2 (25.0)	5 (55.6)	6 (66.7)	0.16
Clinician	1 (5.6)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)	0.82
Clinician educator	8 (44.4)	3 (60.0)	1 (50.0)	2 (40.0)	2 (33.3)	
Clinician scientist	8 (44.4)	2 (40.0)	1 (50.0)	2 (40.0)	3 (50.0)	
Other/I don't know	1 (5.6)	0 (0.0)	0 (0.0)	0 (0.0)	1 (16.7)	
Nonacademic	3 (8.3)	1 (10.0)	0 (0.0)	0 (0.0)	2 (2.22)	
Hybrid/Other	15 (41.7)	4 (40.0)	6 (75.0)	4 (44.4)	1 (11.1)	
Annual colonoscopy volume, n (%)						
<50	–	–	–	–	–	–
50–100	7 (19.4)	2 (20.0)	1 (12.5)	2 (22.2)	2 (22.2)	0.51
101–300	23 (63.9)	7 (70.0)	4 (50.0)	6 (66.7)	6 (66.7)	
301 –500	5 (13.9)	1 (10.0)	3 (37.5)	0 (0.0)	1 (11.1)	
501–700	1 (2.8)	0 (0.0)	0 (0.0)	1 (11.1)	0 (0.0)	
>700	–	–	–	–	–	–
Days per week ofendoscopy, n (%)						
0–1	18 (50.0)	6 (60.0)	3 (37.5)	5 (55.6)	4 (44.4)	0.82
2–3	12 (33.3)	2 (20.0)	4 (50.0)	2 (22.2)	4 (44.4)	
4–5	6 (16.7)	2 (20.0)	1 (12.5)	2 (22.2)	1 (11.1)	
Prior instruction on polyp sizing, n (%)						
Yes	6 (16.7)	3 (30.0)	1 (12.5)	0 (0.0)	2 (22.2)	0.17
No	28 (77.8)	7 (70.0)	7 (87.5)	7 (77.8)	7 (77.8)	
I don't know	2 (5.6)	0 (0.0)	0 (0.0)	2 (22.2)	0 (0.0)	

NA, not applicable.

[1]Chi-squared tests used to compare intervention groups.

Table 2

Confidence level of polyp size classification by group.

	All polyps, n (%)	Diminutive (0–5 mm), n (%)	Small (6–9 mm), n (%)	Large (≥10 mm), n (%)
Control (N = 377)				
High confidence	304 (80.6)	202 (82.1)	65 (76.5)	37 (80.4)
Low confidence	73 (19.4)	44 (17.9)	20 (23.5)	9 (19.6)
Video only (N = 297)				
High confidence	222 (74.7)	156 (80.4)	43 (64.2)	23 (63.9)
Low confidence	75 (25.3)	38 (19.6)	24 (35.8)	13 (36.1)
Feedback only (N = 326)				
High confidence	243 (74.5)	164 (77.7)	46 (62.2)	33 (80.5)
Low confidence	83 (25.5)	47 (22.3)	28 (37.8)	8 (19.5)
Video + Feedback (N = 360)				
High confidence	290 (80.6)	205 (87.6)	56 (69.1)	29 (64.4)
Low confidence	70 (19.4)	29 (12.4)	25 (30.9)	16 (35.6)
Overall (N = 1360)				
High confidence	1059 (77.9)	727 (82.1)	210 (68.4)	122 (72.6)
Low confidence	301 (22.1)	158 (17.9)	97 (31.6)	46 (27.4)