The Paddington International Virtual Chromoendoscopy Score in ulcerative colitis exhibits very good inter-rater agreement after computerized module training:

Trivedi, Palak; Kiesslich, Ralf; Hodson, James; Bhala, Neeraj; Boulton, Ralph; Cooney, Rachel; Gui, Xianyong; Iqbal, Tariq; Li, Ka-Kit; Mumtaz, Saqib; Pathmakanthan, Shri; Quraishi, Mohammed Nabil; Sagar, Vandana; Shah, Ashit; Sharma, Naveen; Siau, Keith; Smith, Samuel; Ward, Stephen; Widlak, Monika M.; Bisschops, Raf

DOI:
10.1016/j.gie.2018.02.044

The Paddington International virtual ChromoendoScopy ScOre (PICaSSO) in ulcerative colitis exhibits very good inter-rater agreement after computerised module training: a multi-centre study across academic and community practice.

Palak J. Trivedi,[1,2,3,4] Ralf Kiesslich,[5] James Hodson,[4] Neeraj Bhala,[3] Ralph A. Boulton,[3] Rachel Cooney,[3] Xianyong Gui,[6] Tariq Iqbal,[3] Ka-kit Li,[7] Saqib Mumtaz,[8] Shri Pathmakanthan,[3] Mohammed N. Quraishi,[3] Vandana M. Sagar,[1,2] Ashit Shah,[8] Naveen Sharma,[9] Keith Siau,[8] Samuel Smith,[3] Stephen Ward,[10] Monika M. Widlak,[11,12] Raf Bisschops,[13] Subrata Ghosh [3,4,14] and Marietta Iacucci.[3,4,14]

1. National Institute of Health Research (NIHR) Birmingham, Liver Biomedical Research Centre (BRC), University of Birmingham, United Kingdom.
2. Liver Unit, University Hospitals Birmingham Queen Elizabeth, United Kingdom
3. Department of Gastroenterology, University Hospitals Birmingham Queen Elizabeth, United Kingdom
4. Institute of Translational of Medicine, Institute of immunology and immunotherapy and NIHR Biomedical Research Centre, University of Birmingham, United Kingdom
5. Department of Medicine, Division of Gastroenterology, HSK Hospital, Wiesbaden, Germany
6. Department of Pathology and Laboratory Medicine, University of Calgary and Calgary Laboratory Services, Canada

7. Department of Gastroenterology and Hepatology, Leicester Royal Infirmary, United Kingdom
8. Department of Gastroenterology, Royal Wolverhampton Hospitals NHS Trust, United Kingdom
9. Department of Gastroenterology, University Hospitals Birmingham Heart of England Foundation Trust, United Kingdom
10. Department of Colorectal Surgery, University Hospitals Coventry and Warwickshire, United Kingdom
11. Department of Gastroenterology, University Hospitals Coventry and Warwickshire, United Kingdom
12. Warwick Medical School, University of Warwick, United Kingdom
13. Department of Gastroenterology and Hepatology, University Hospitals Leuven, KU Leuven, Leuven, Belgium
14. Division of Gastroenterology, University of Calgary, Alberta Canada

Correspondence:

Marietta Iacucci MD, PhD

Reader/Senior Associate Professor of Gastroenterology

Institute of Translational Medicine

University of Birmingham

Adjunct Clinical Associate Professor of Medicine

University of Calgary

Institute of Translational Medicine

Heritage Building Research & Development

University Hospital Birmingham NHS Foundation Trust

Edgbaston B15 2TT Birmingham, UK

Telephone +44(0)1213718119

email : m.iacucci@bham.ac.uk

Conflicts of Interest and funding disclosures

MI: received unrestricted research grant from Pentax USA

PJT: received institutional salary support from the NIHR.

None of the other authors have any conflict of interest to declare

This article is independent research supported by the NIHR Birmingham Liver Biomedical Research Centre. The views expressed in this publication are of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health.

MI is funded by the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflicts of Interest

Author contributions:

Study design and concept: MI, SG

Data Acquisition: PJT, NB, RAB, RC, TI, KkL, SM, SP, MNQ, VMS, AS, NS, KS, SS, SW, MMW

Analysis of data: PJT, JH

Writing and preparation of manuscript: PJT, MI, SG

Finalising of manuscript to completion: PJT, NB, RAB, RB, RC, XG, JH, TI, RK, KkL, SM, SP, MNQ, VMS, AS, NS, KS, SS, SW, MMW, SG, MI

Abstract

Introduction: Electronic virtual chromoendoscopy (EVC) can demonstrate ongoing disease activity in ulcerative colitis (UC) even when Mayo subscores suggest healing. However, applicability of EVC technology outside the expert setting has yet to be determined.

Methods: 15 participants across five centres reviewed a computerised training module outlining high definition (HD) and EVC (i-Scan) colonoscopy modes. Inter-observer agreement was then tested (Mayo score, UCEIS and the **P**addington **I**nternational virtual **C**hromoendo**S**copy **S**c**O**re [PICaSSO] for UC), using a colonoscopy video library (n=30 cases reviewed pre- and n=30 post-training). Knowledge sustainability was re-tested in a second round (n=42 cases; 9/15 participants), 6 months post-training provision.

Results: Pre-training intraclass correlation coefficients (ICC) were good for the Mayo endoscopic subscore (ICC:0.775), UCEIS scoring erosions/ulcers (ICC:0.770) and UCEIS overall (ICC:0.786), and for mucosal (ICC:0.754) and vascular components of PICaSSO (ICC:0.622). For the vascular components of UCEIS, agreement was only moderate (ICC:0.429), and did not enhance post-training (ICC:0.417); unlike for PICaSSO which improved (mucosal ICC:0.848; vascular: 0.746). Histological correlation using the New York Mt. Sinai System was strong for both PICaSSO components (Spearman's *rho* for mucosal: 0.925, and vascular: 0.873; $p<0.001$ for both). Moreover, accuracy in specifically discriminating quiescent from mild histological strata was strongest for PICaSSO (AUROC for mucosal: 0.781; vascular: 0.715), compared to Mayo (AUROC:0.708) and UCEIS (AUROC for UCEIS overall: 0.705; vascular: 0.562; bleeding: 0.645; erosions/ulcers: 0.696). Inter-rater reliability for PICaSSO was sustained

by round two participants (Round 1 and 2 ICC for mucosal: 0.873 and 0.869, respectively; and vascular: 0.715 and 0.783, respectively), together with histological correlation (*rho* mucosal: 0.934, vascular: 0.938; $p<0.001$ for both).

Conclusion: PICaSSO demonstrates good inter-observer agreement across all levels of experience, providing excellent correlation with histology. Given ability to discriminate subtle endoscopic features, PICaSSO may be applied to refine stratified treatment paradigms for UC patients.

Keywords: Colonoscopy, endoscopic remission, histological remission, inflammatory bowel disease, mucosal healing, risk stratification

Study Highlights

What is already known on the issue?

Conventional white light endoscopy have limitations in defining inflammation in ulcerative colitis, especially at the milder end of the spectrum and mucosal healing. More detailed assessment such as histologic scoring may better predict relapses. Electronic virtual chromoendoscopy (EVC) can demonstrate ongoing disease activity even when Mayo scores suggest healing. EVC score (PICaSSO) was designed and validated by international expert endoscopists in EVC. However, applicability of EVC scoring in UC requires validation outside the expert setting.

How was the study done?

A training module was first developed. Fifteen participants across five general gastroenterology and colorectal divisions reviewed a computerised training module outlining high definition (HD) and EVC (i-Scan) colonoscopy modes. Inter-observer agreement was then tested (Mayo score, UCEIS and the **P**addington **I**nternational virtual **C**hromoendo**S**copy **S**c**O**re [PICaSSO] for UC), using a colonoscopy video library (n=30 cases reviewed pre- and n=30 post-training). An abbreviated and simple PICaSSO score (PICaSSO-ab) was created. Knowledge sustainability was re-tested in a second round (n=42 cases; 9/15 participants), 6 months post-training provision.

What were the main findings?

Agreement for vascular components of PICaSSO was good (ICC 0.622) and better than for the vascular components of UCEIS, where agreement was only moderate (ICC:0.429), and did not enhance post-training (ICC:0.417). This was unlike PICaSSO which improved (mucosal ICC: 0.848; vascular: 0.746). Histological correlation using the New York Mt. Sinai System was strong for both PICaSSO components (Spearman's *rho* for mucosal: 0.925, and vascular: 0.873; $p<0.001$ for both). Moreover, accuracy in specifically discriminating quiescent from mild histological strata was strongest for PICaSSO (AUROC for mucosal: 0.781; vascular: 0.715), compared to Mayo (AUROC:0.708) and UCEIS (AUROC for UCEIS overall: 0.705; vascular: 0.562; bleeding: 0.645; erosions/ulcers: 0.696). Inter-rater reliability for PICaSSO was sustained in round two participants together with histological correlation (*rho* mucosal: 0.934, vascular: 0.938; $p<0.001$ for both).

What do these results add to the current body of knowledge?

The Paddington International virtual ChromoendoScopy ScOre (PICaSSO) achieves good inter-rater reliability post-training, across all levels of endoscopy experience. Correlation between PICaSSO and histology is strong, with performance accuracy that is sustainable over time. PICaSSO provides the most accurate discrimination between quiescent and mild histological disease activity, compared to the Mayo score and UCEIS.

Abbreviations

AUROC:	Area under the receiver operating characteristic curve
AGA:	American Gastroenterogical Association
BLI:	Blue Laser Imaging
EVC:	Electronic virtual chromoendoscopy
HD:	High definition
OE:	Optical enhancement
ICC:	Intraclass correlation coefficient
MH:	Mucosal healing
NBI:	Narrowband Imaging
PICaSSO:	Paddington International virtual ChromoendoScopy ScOre
ROC:	Receiver operating characteristic
UC:	Ulcerative colitis
UCEIS:	Ulcerative Colitis Endoscopic Index of Severity
UK:	United Kingdom

WLE: White light endoscopy

INTRODUCTION

Treatment paradigms in ulcerative colitis (UC) have been revolutionised by the advent of novel, targeted therapies;[1–6] with mucosal healing (MH) a critical endpoint for clinical trials and in practice.[7–9] Consequently, many endoscopic indices have been devised in UC,[7,8,10] with the Mayo endoscopic subscore the most widely adopted.[11] Despite its popularity, the Mayo score has been criticised for including descriptors that overlap between different tiers, inclusion of terms open to subjective interpretation (e.g. friability), multiple descriptors within same scoring tiers, lack of clear definition of mucosal healing and wide inter-observer variability. Additionally, clinical trials in UC often classify Mayo 0–1 collectively as MH,[1,12] despite significant differences in the incidence of disease relapse and colectomy between the two groups.[13,14] Furthermore, abnormalities in fine vascular pattern rather than definitive loss, is generally apparent with usage of contemporary high resolution colonoscopy; with the Mayo endoscopic subscore largely representing t legacy of older generation endoscopes or even rigid sigmoidoscopy. More contemporary scoring systems have been proposed, most notably the validated Ulcerative Colitis Endoscopic Index of Severity (UCEIS),[15,16] . The UCEIS demonstrates good correlation with disease severity and prediction of clinical outcome by experts in the field, [17–19] but does not define the endoscopic features of MH in UC. UCEIS also categorises 'loss' of vascular pattern, rather than describing the fine vascular abnormalities visible at high resolution.

A notable caveat of existing endoscopic scoring systems is that all were derived using conventional white light endoscopy (WLE). Data from the Oxford group illustrates that

>20% of patients exhibit persistent inflammation histologically, even after attaining endoscopic remission as determined by WLE.[20] Indeed, histological remission is a target distinct from endoscopic mucosal healing, and may better predict the incidence of future clinical events.[20] However this may reflect the older generation of endoscopes with lower resolution capabilities. In a similar vein, it has been shown that 30% of individuals having a Mayo endoscopic subscore of 0 exhibit persistent inflammatory mucosal changes when re-examined with high definition (HD) electronic virtual chromoendoscopy (EVC).[21] This has furthered development of a dedicated EVC scoring system to better assess inflammation, and quantify how the measure of abnormal vascular pattern correlates with histology.[22] However, there is clear need for effective and validated training in EVC technology; and computer or web-based teaching may offer an opportunity for easy, and inexpensive delivery. Indeed, with new technology the paradigms of endoscopic classification in UC are changing.

The **P**addington **I**nternational virtual **C**hromoendo**S**copy **S**c**O**re (PICaSSO), developed and validated following provision of a comprehensive computerised training module, represents the first EVC-based assessment tool for systematically evaluating disease activity in UC.[22] Despite good inter-observer agreement between raters and strong correlation with histological indices, a limitation to the original study is that all participants were dedicated endoscopy experts; whereas operating characteristics in the gastroenterology community as a whole has not been determined. Thus, the principal aim of this study is to validate inter-rater reliability of PICaSSO specifically in a non-expert setting, across a breadth of endoscopy experience and to assess the sustainability of diagnostic performance over time after training. We also attempted to simplify PICaSSO

for use by those less experienced in EVC assessment, and develop a standardised training module that may be adopted in wider clinical practice.

METHODS

We tested the external reproducibility of PICaSSO in experienced consultants and trainees, who had no prior exposure to EVC iSCAN virtual Chromoendoscopy. Investigators for this study comprised practising gastroenterologists and colorectal surgeons from five United Kingdom (UK) hospitals (**Figure 1**).

Training module design

One gastroenterologist (MI) with expertise in EVC and optical diagnosis in IBD developed a training module based on pre-existing UC scoring systems and PICaSSO;[11,15,22] encompassing the following characteristics:

- The clinical importance in differentiation between quiescent vs. mild activity in UC.
- Limitations of Mayo endoscopic scoring and UCEIS.
- The PICaSSO EVC score and detailed characterisation of mucosal and vascular changes.

MI did not score videos in the study.

The training module consisted of 100 high-resolution endoscopic pictures and 10 videos, and was assessed by all the participants with direct feedback and stepwise discussion. These illustrated the entire spectrum of inflammatory mucosal and vascular changes

including mucosal healing in UC (**Figure 2** and **Supplementary Video 1** & VCE PICaSSO training module with a link to it via website)

A. Round one

Pre-training

All participants were presented with 30 colonoscopy videos (quiescent, *n* = 10; mild, *n* = 7; moderate, *n* = 6; severe, *n* = 7 by histology grading) recorded in high definition white light and i-SCAN modes (EPKi 7000 Pentax EC-3490Fi in 3 settings, high-definition white light i-SCAN 1 and EVC i-SCAN 2 and 3 modes integrated into the endoscope hand piece that can be operated by simply pressing on the button),. Each rater scored individual videos: using a standardised case record form as indicated previously,[22] prior to any training provision (pre-training component). Participants were also provided with a printed sheet, listing the individual anchor points of the Mayo score, UCEIS and PICaSSO; but with no illustrations/photographic material shown.

Training module provision

Thereafter, the pre-designed 60-minute comprehensive training module was delivered (by MI and SG), including an introduction to the study, explanation of endoscopic scores in use, and detailing all the endoscopic findings (different selection from the pre-/post-training video library) of varying grades of inflammation in UC; illustrated via HD and EVC modes (**Figure 2, Supplementary EVC PICaSSO Training module** and **Table 1A**).

Post-training

Following training, each participant scored the 30 videos in a different random order (post-training component) and the results compared. All videos (in the pre- and post-training modules) were scored according to the Mayo endoscopic subscore,[11] UCEIS[15,16] and PICASSO.[22] In addition to the original, highly detailed scoring system proposed by Iacucci *et al.*,[22] the inter-rater reliability of an abbreviated, simpler version of PICaSSO was also tested (PICaSSO-Ab.) (**Table 1B**). The original PICaSSO score was designed with items that had the potential to be simplified based on the operating characteristics in the original validation.[22] The abbreviated PICaSSO score was devised based on the results of the multivariate analysis of the endoscopic predictors of the grade of inflammation at histology, done item by item. The endoscopic subcategories of mucosal or intraluminal bleeding with round dilated or crowded tortuous superficial vessels as well as erosions or ulcerations, discrete or diffuse, were not important predictors of the multivariate analysis of grade of inflammation at histology.

B. Round two

In the second validation phase we assessed the long-term learning, sustainability and diagnostic performance of the PICaSSO endoscopic scoring system in UC. All participants were invited to attend a second round after a six-month interval (during which no exposure to PICaSSO scoring and i-SCAN technology took place). Each individual scored a new set of 42 videos (quiescent, n = 15; mild, n = 8; moderate, n = 6; severe, n = 13), which were different from the 30 videos previously assessed.

Video library selection for both rounds

A total of 72 unique, high-quality video clips (30 + 42) representing collections of different grades of inflammation were selected in the first and second phases of the study from an existing library. This anonymised library was collated by one investigator (MI) from colonoscopies assessing the breadth of inflammatory activity observed in UC patients (**Figure 2**). Videos were saved in Audio Video Interleave (AVI) format (S-video output to endoPRO legacy, MPS Motion Picture Studio; standard definition image capture in MPEG3).

A single pathologist (XG) blinded to the results of endoscopic scoring scored the histological severity of disease for each video, according to the New York Mt. Sinai System as proposed by Hefti *et al*.[23]

Data interpretation and analysis

Given the large number of participants and a desire to maintain the ordinal hierarchy of scoring, agreement between participants was tested using intra-class correlation coefficients (ICC).[26] Results of ICC analysis were classified as very good (coefficients: 0.81 – 1.00), good (0.61 – 0.80), moderate (0.41 – 0.60), fair (0.21 – 0.40) and poor (<0.21).[24] In addition to the ICCs, a 'pairs of raters' approach was also used. This compared the scores given every possible pair of participants, and was reported as the proportion of those that matched. The resulting statistic represented the probability that two randomly selected participants would give the same exact score for a given patient.

Correlations between the individual endoscopic scores, averaged across all participants, and histological indices of disease severity were assessed using Spearman's Rank

correlation coefficients (rho). All analyses were performed using IBM SPSS 22 (IBM Corp. Armonk, NY), with $p<0.05$ deemed to be indicative of statistical significance throughout.

RESULTS

Participants

Participants for this study in round one comprised practicing gastroenterologists and colorectal surgeons at consultant (n = 7) and trainee level (n = 8) from five UK hospitals. Lifetime procedure counts varied between 200 colonoscopies for the most junior participant (range 200 – 450), to 4,500 for the most senior (range 2,000 – 4,500). Whilst participants indicated familiarity with the Mayo endoscopic subscore and UCEIS, and narrowband imaging (NBI) endoscopic system, none were experienced in i-SCAN EVC technology.

Inter-observer reliability of endoscopic scores following computerised module training

The inter-rater agreement was good for the Mayo score, the UCEIS score (as well as the erosions/ulcers and bleeding components), and for both the PICaSSO mucosal and vascular patterns in the pre-training module (**Table 2**). Inter-rater agreement was weakest for the vascular component of UCEIS, for which agreement was only moderate.

After training, inter-rater agreement improved for the majority of scores considered, with the Mayo score, the overall UCEIS score (as well as the individual erosion/ulcer component) and the PICaSSO vascular pattern now achieving ‘very good’ agreement

(ICC>0.8). The only score not to improve with training was the vascular component of UCEIS, for which agreement remained moderate.

Correlation of the PICaSSO EVC score with histological indices

In the data collected post-training, strong correlations were detected between the histological score and PICaSSO for both the full and abbreviated (PICaSSO-ab) versions of the mucosal (Spearman's *rho:* 0.925 [95% CI: 0.843 – 0.965], $p<0.0001$; and 0.894 [0.783 – 0.950], $p<0.0001$; respectively) and vascular components (*rho:* 0.873 [0.743 – 0.940], $p<0.0001$; and 0.889 [0.772 – 0.947], $p<0.0001$; respectively) (**Figure 3**). Similar correlations were also detected between the histological score and both the UCEIS and the Mayo endoscopic subscores (UCEIS vascular component: 0.829 [0.662 – 0.918], bleeding component: 0.913 [0.819 – 0.959], mucosal component: 0.872 [0.742 – 0.939], overall: 0.887 [0.770 – 0.947]; Mayo endoscopic subscore: 0.876 [0.748 – 0.941]; *p* values <0.0001 for all).

Inter-rater reliability across histological strata

Next, the degree of reliability across each histological strata was determined. By analysing every possible combination of rater scores (n = 15 participants; 105 possible rater combinations), we found that agreement was stronger for PICaSSO-Ab. vs. the full PICaSSO system across both components, being greatest at extremes of disease activity (**Figure 4**) (**Supplementary Table 1**), though of course PICaSSO-Ab lost some of the details of the full PICaSSO. When observing the percentage agreement across all tested

scoring systems, it was evident that mucosal components in PICaSSO and UCEIS had consistently better inter-rater reliability than their vascular/bleeding counterparts.

Accuracy in predicting quiescent disease

A Mayo score of 0 is commonly applied as an endoscopic endpoint in clinical trials, although histological disease activity may yet persist and better forecast outcomes for UC patients.[20,21] Thus, in an effort to better discriminate quiescent vs. mild histological disease activity we conducted sensitivity analysis for each endoscopic scoring system. Evaluating all individual rater responses on a per-video basis (n = 450 data points), the most accurate scoring system predictive of quiescent disease was the mucosal component of PICaSSO (area under the receiver operator characteristic curve [AUROC]: 0.917, 95% CI: 0.890 – 0.943), $p < 0.001$; **Figure 5a**). Moreover, in a restricted analysis of only those cases capturing quiescent vs. mild histological disease activity (n = 255 data points), the highest performing scoring system was PICaSSO across both components (**Figure 5b**) (**Supplementary Table 3**).

Sustainability of performance of PICaSSO over time

A second round of video scoring took place following a 6-month interval, in which participants were invited to score a further 42 videos. Of the original fifteen raters, 9 responded to the invitation and participated in round two (5 consultants and 4 trainees).

No participant was exposed to iscan endoscopic system (or other EVC technology) to assess UC during this period. The level of inter-rater reliability in this group at the second session was found to be consistent with that observed in the post-training assessment in round one (results for round one and round two for the 9/15 participants are provided in

Table 3). The same was true for the pairs-of-raters approach, broken down by histological strata (**Supplementary Table 2**).

Correlations between the PICaSSO score and histological strata also remained strong for both the full (Spearman's *rho* for mucosal component: 0.934 [0.878 – 0.965]; vascular component: 0.938 [0.885 – 0.967]; $p<0.0001$ for both) and abbreviated systems (mucosal component: 0.927 [0.867 – 0.961]; vascular component: 0.909 [0.836 – 0.951], respectively; $p<0.0001$ for both).

DISCUSSION

We show that the newly developed EVC score for UC, PICaSSO, detailing mucosal and vascular components through HD i-SCAN technology, is easy to learn, displays very good reliability between raters at all levels of endoscopy training after provision of a computerised training module. We further demonstrate that the diagnostic performance of PICaSSO was sustained over time (six months); and that the predictive accuracy is greater than the Mayo score and UCEIS in permitting discrimination between quiescent vs. mild histological disease activity. To our knowledge, this is the first effective and standardised training module that can be applied across all levels of endoscopy experience, which allows EVC scoring modalities to be implemented in clinical practice.

Accurate assessment of disease activity is critical for guiding treatment decisions in UC, both with regard to escalating and deescalating therapy, as well as recognition of dysplastic change. In this regard, endoscopic image-enhancement allows visualisation of

normal and abnormal mucosa complementary to conventional white light endoscopic imaging.[10] During colonoscopic surveillance, EVC is also proven to better lesion recognition and characterisation, and facilitate precision-targeted biopsies to a greater degree than standard white-light endoscopy.[25] However, despite widespread availability for over a decade, utilising EVC beyond academic research has been limited by lack of standardised training and paucity of guideline-directed implementation into routine clinical practice. In the meantime, EVC technology has progressed and been refined rapidly.

The ability to train endoscopists without experience in EVC technology to reach acceptable levels of competence has been the focus of the recent white paper by the American Gastroenterological Association (AGA).[26] Therein, it was acknowledged that computer-based training modules with ongoing reflective feedback are an effective 'tool for training,' particularly with reference to lesion recognition using NBI. Along similar lines, our study represents the first external validation of an EVC scoring system assessing ulcerative colitis disease activity across a breadth of colonoscopy experience. Importantly, the pool of endoscopists who participated all denied familiarity with i-SCAN technology. Through provision of a computerised training module, we demonstrate very good inter-rater reliability of PICaSSO and strong correlation with histological activity, both of which are sustained over time. This suggests that the learning curve for PICaSSO is short, effective, and accomplishable within several hours of concentrated training. The testament to our exercise is perhaps best highlighted by the fact that for all tested scoring systems (except the UCEIS component of vascular change), the degree of reliability improved following training provision. Moreover, we were in line with other recent studies

independently evaluating the usefulness of brief training interventions in improving diagnostic accuracy of colonic polyp recognition for clinicians with varying endoscopic expertise.[27–30]

Preliminary data using magnifying NBI of colonic mucosa suggests that UC relapse may be predicted by vascular changes alone, without considering the mucosal findings.[31] The vascular features captured by NBI as well as BLI are similar to i-SCAN, and it is probable that a common EVC score may be developed in future irrespective of the chosen platform; a concept we hope to develop as a prospective multicenter endeavour which is now ongoing. Indeed, endoscopic features of mucosal and vascular healing may be defined across all EVC platforms (iSCAN , NBI and BLI) with scoring tiers very similar to the PICaSSO score. We are evaluating this in a multicentre setting. To this effect we have provided matching images of different grades of activity for NBI and BLI to indicate how translation to other endoscopic EVC platforms may be feasible. (Figure 6,7,8) The endoscopic findings of mucosal and vascular healing – such as continuous/regular crypts, crypts not visible (scar), discontinuous and or dilated/elongated crypts; and vascular changes such as 'roundish' appearance following crypt architecture, vessels not visible (scar), and sparse deep vessels without dilatation – are individually and collectively shown to predict histological remission.[22] Their critical importance of these concepts is perhaps best highlighted when discriminating quiescent vs. mild histological disease activity in sensitivity analysis, wherein both PICaSSO components performed best across all scoring systems tested; potentially impacting real-time treatment decisions and stratified patient care.

Early studies with other scoring systems have shown a large variance in agreement between gastroenterologists to assess the endoscopic activity in UC.[32] However, the intraclass correlation coefficients we obtained for both PICaSSO components were comparable to the Mayo endoscopic subscore and UCEIS; both internally and versus results of central readership in certain clinical trials.[33] Of note, conventional methods of testing inter-rater reliability that only apply kappa statistics do not allow accurate weighting, and suffer from the issue of joint-probability by treating data as nominal rather than ordinal in nature. For instance, the variability in scoring between luminal bleeding and vessel dilatation is clearly greater than the clinical divide between vessel dilation and vascular crowding, which is captured by ICC methodology but not via Kappa statistics. Whilst Cohen's weighted Kappa may allow a degree of hierarchy to be maintained, this is not possible to apply in our study given the large number of individual raters. Intraclass correlation coefficients are also 'chance-corrected,' resulting in a more accurate measure of agreement and not affected by the different number of components in any given score.[34]

One of the pre-requisites for ensuring completeness to our video library, both in the training module and for test scoring, was to encompass the full spectrum of inflammatory activity observed in ulcerative colitis. In so doing, we found that study participants performed better in assessing mucosal lesions as opposed to vascular patterns, akin to the expert panel that comprised the original London Paddington consensus.[22] This can partially be explained by the fact that EVC is a relatively new concept in UC, although the vascular component of the UCEIS also under performed. Thus, we need to continually improve the vascular component both for definition and for training purposes.

In this study, an abbreviated version of PICaSSO was also proposed, given that early mucosal and vascular changes exclude the presence of erosions/ulcers and onset of bleeding, respectively; hence subtlety and detail in scoring carries less relevance during overt, active disease states. Despite this restriction, inter-rater reliability was sustained, durable over time, and exhibited strong parallels with histological correlates. Moreover, PICaSSO-Ab. system, for both mucosal and vascular components, improved the level of agreement across individual histological strata. The advantages of PICaSSO-Ab. are mirrored in its simplicity; although further in-depth learning curve assessment is critical, with an extended teaching phase as proposed by the original investigators.[22]

This study has several limitations . The same videos have been used in the pre and post test in round 1 which was performed in the same day. There is a possibility of recall bias despite the videos had been scored in randomized different order. It is also difficult to recall PICaSSO..In any event , the K interobserver agreement between raters was good and sustained after 6 months where in a completely new set of videoclips were scored.

In addition, the PiCaSSO score requires validation in real time clinical practice ,and correlation with outcomes that are clinically meaningful; specifically relapses in active disease, the frequency of hospitalisation episodes and time to colectomy especially within of patients exhibiting endoscopic Mayo Scores of 0 and 1. Further work is also ongoing to converge single yet comprehensive and accurate scoring system applicable across all EVC platforms.

In conclusion, PICaSSO displays very good and sustained reliability between raters, both experienced and less experienced endoscopists, after delivery of a standardised training module with good correlation against histological gradings. Future endeavours will need to evaluate the performance of PICaSSO across a multicentre prospective setting in real time (currently underway), in addition to assessing how clinical outcome prediction can be enhanced through EVC technology in UC.[19,26]

References

1. Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, Johanns J, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2005 Dec 8;353(23):2462–76.
2. Feagan BG, Rutgeerts P, Sands BE, Hanauer S, Colombel J-F, Sandborn WJ, et al. Vedolizumab as induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2013 Aug 22;369(8):699–710.
3. Sandborn WJ, Feagan BG, Stoinov S, Honiball PJ, Rutgeerts P, Mason D, et al. Certolizumab pegol for the treatment of Crohn's disease. N Engl J Med. 2007 Jul 19;357(3):228–38.
4. Sandborn WJ, Feagan BG, Wolf DC, D'Haens G, Vermeire S, Hanauer SB, et al. Ozanimod Induction and Maintenance Treatment for Ulcerative Colitis. N Engl J Med. 2016 May 5;374(18):1754–62.
5. Vermeire S, O'Byrne S, Keir M, Williams M, Lu TT, Mansfield JC, et al. Etrolizumab as induction therapy for ulcerative colitis: a randomised, controlled, phase 2 trial. Lancet. 2014 Jul 26;384(9940):309–18.
6. Sandborn WJ, Su C, Sands BE, D'Haens GR, Vermeire S, Schreiber S, et al. Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med. 2017 04;376(18):1723–36.
7. Walsh AJ, Bryant RV, Travis SPL. Current best practice for disease activity assessment in IBD. Nat Rev Gastroenterol Hepatol. 2016 Oct;13(10):567–79.
8. Samaan MA, Mosli MH, Sandborn WJ, Feagan BG, D'Haens GR, Dubcenco E, et al. A systematic review of the measurement of endoscopic healing in ulcerative

colitis clinical trials: recommendations and implications for future research. Inflamm Bowel Dis. 2014 Aug;20(8):1465–71.

9. Vuitton L, Peyrin-Biroulet L, Colombel JF, Pariente B, Pineton de Chambrun G, Walsh AJ, et al. Defining endoscopic response and remission in ulcerative colitis clinical trials: an international consensus. Aliment Pharmacol Ther. 2017 Mar 1;45(6):801–13.
10. Iacucci M, Panaccione R. Recent advances in the endoscopic assessment of ulcerative colitis. Tech Gastrointest Endosc. 2016 Jul 1;18(3):116–22.
11. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med. 1987 Dec 24;317(26):1625–9.
12. Cooney RM, Warren BF, Altman DG, Abreu MT, Travis SP. Outcome measurement in clinical trials for Ulcerative Colitis: towards standardisation. Trials. 2007 Jun 25;8:17.
13. Barreiro-de Acosta M, Vallejo N, de la Iglesia D, Uribarri L, Bastón I, Ferreiro-Iglesias R, et al. Evaluation of the Risk of Relapse in Ulcerative Colitis According to the Degree of Mucosal Healing (Mayo 0 vs 1): A Longitudinal Cohort Study. J Crohns Colitis. 2016 Jan 1;10(1):13–9.
14. Manginot C, Baumann C, Peyrin-Biroulet L. An endoscopic Mayo score of 0 is associated with a lower risk of colectomy than a score of 1 in ulcerative colitis. Gut. 2014 Dec 30;gutjnl-2014-308839.
15. Travis SPL, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel J-F, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the

Ulcerative Colitis Endoscopic Index of Severity (UCEIS). Gut. 2012 Apr;61(4):535–42.

16. Travis SPL, Schnell D, Krzeski P, Abreu MT, Altman DG, Colombel J, et al. Reliability and Initial Validation of the Ulcerative Colitis Endoscopic Index of Severity. Gastroenterology. 2013 Nov 1;145(5):987–95.
17. Arai M, Naganuma M, Sugimoto S, Kiyohara H, Ono K, Mori K, et al. The Ulcerative Colitis Endoscopic Index of Severity is Useful to Predict Medium- to Long-Term Prognosis in Ulcerative Colitis Patients with Clinical Remission. J Crohns Colitis. 2016 Nov;10(11):1303–9.
18. Saigusa K, Matsuoka K, Sugimoto S, Arai M, Kiyohara H, Takeshita K, et al. Ulcerative colitis endoscopic index of severity is associated with long-term prognosis in ulcerative colitis patients treated with infliximab. Dig Endosc. 2016 Sep;28(6):665–70.
19. Ikeya K, Hanai H, Sugimoto K, Osawa S, Kawasaki S, Iida T, et al. The Ulcerative Colitis Endoscopic Index of Severity More Accurately Reflects Clinical Outcomes and Long-term Prognosis than the Mayo Endoscopic Score. J Crohns Colitis. 2016 Mar;10(3):286–95.
20. Bryant RV, Burger DC, Delo J, Walsh AJ, Thomas S, von Herbay A, et al. Beyond endoscopic mucosal healing in UC: histological remission better predicts corticosteroid use and hospitalisation over 6 years of follow-up. Gut. 2016 Mar;65(3):408–14.
21. Iacucci M, Fort Gasia M, Hassan C, Panaccione R, Kaplan GG, Ghosh S, et al. Complete mucosal healing defined by endoscopic Mayo subscore still demonstrates

abnormalities by novel high definition colonoscopy and refined histological gradings. Endoscopy. 2015 Aug;47(8):726–34.

22. Iacucci M, Daperno M, Lazarev M, Arsenascu R, Tontini GE, Akinola O, et al. Development and reliability of the new endoscopic virtual chromoendoscopy score: the PICaSSO (Paddington International Virtual ChromoendoScopy ScOre) in ulcerative colitis. Gastrointest Endosc [Internet]. 2017; Available from: http://www.sciencedirect.com/science/article/pii/S001651071730192X

23. Hefti MM, Chessin DB, Harpaz N, Steinhagen RM, Ullman TA. Severity of Inflammation as a Predictor of Colectomy in Patients With Chronic Ulcerative Colitis. Dis Colon Rectum. 2009 Feb;52(2):193–7.

24. Oremus M, Oremus C, Hall GBC, McKinnon MC, &amp ECT, Team CSR. Inter-rater and test–retest reliability of quality assessments by novice student raters using the Jadad and Newcastle–Ottawa Scales. BMJ Open. 2012 Jan 1;2(4):e001368.

25. Gasia MF, Ghosh S, Panaccione R, Ferraz JG, Kaplan GG, Leung Y, et al. Targeted Biopsies Identify Larger Proportions of Patients With Colonic Neoplasia Undergoing High-Definition Colonoscopy, Dye Chromoendoscopy, or Electronic Virtual Chromoendoscopy. Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc. 2016 May;14(5):704–712.e4.

26. Gupta N, Brill JV, Canto M, DeMarco D, Fennerty BM, Laine L, et al. AGA White Paper: Training and Implementation of Endoscopic Image Enhancement Technologies. Clin Gastroenterol Hepatol. 2017 Jun;15(6):820–6.

27. Rastogi A, Rao DS, Gupta N, Grisolano SW, Buckles DC, Sidorenko E, et al. Impact of a computer-based teaching module on characterization of diminutive colon polyps by using narrow-band imaging by non-experts in academic and

community practice: a video-based study. Gastrointest Endosc. 2014 Mar 1;79(3):390–8.

28. Ignjatovic A, Thomas-Gibson S, East JE, Haycock A, Bassett P, Bhandari P, et al. Development and validation of a training module on the use of narrow-band imaging in differentiation of small adenomas from hyperplastic colorectal polyps. Gastrointest Endosc. 2011 Jan;73(1):128–33.

29. Raghavendra M, Hewett DG, Rex DK. Differentiating adenomas from hyperplastic colorectal polyps: narrow-band imaging can be learned in 20 minutes. Gastrointest Endosc. 2010 Sep;72(3):572–6.

30. Patel SG, Schoenfeld P, Kim HM, Ward EK, Bansal A, Kim Y, et al. Real-Time Characterization of Diminutive Colorectal Polyp Histology Using Narrow-Band Imaging: Implications for the Resect and Discard Strategy. Gastroenterology. 2016 Feb;150(2):406–18.

31. Hayashi S, Kudo S, Ogata N, Ohtsuka K, Wakamura K, Maeda Y, et al. Narrow-Band Imaging Efficiency for Evaluation of Mucosal Healing/ Relapse of Ulcerative colitis. Gastrointest Endosc. 2016 May 1;83(5):AB154.

32. Osada T, Ohkusa T, Yokoyama T, Shibuya T, Sakamoto N, Beppu K, et al. Comparison of several activity indices for the evaluation of endoscopic activity in UC: inter- and intraobserver consistency. Inflamm Bowel Dis. 2010 Feb;16(2):192–7.

33. Feagan BG, Sandborn WJ, D'Haens G, Pola S, McDonald JWD, Rutgeerts P, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. Gastroenterology. 2013 Jul;145(1):149–157.e2.

34. Mandrekar JN. Measures of Interrater Agreement. J Thorac Oncol. 2011 Jan 1;6(1):6–7.

Table 1: Components of the PICaSSO EVC score in ulcerative colitis (full and abbreviated)

Mucosal architecture	Vascular architecture
A) PICaSSO (*full version)*	
0. No mucosal defect	***0. Vessels without dilatation***
A) Continuous / regular crypt architecture B) Crypts not visible (scar) C) Discontinuous and/or dilated/elongated crypts	A) Roundish; following crypt architecture B) Vessels not visible (scar) C) Sparse (deep) vessels without dilatation
I. Micro-erosions or crypt abscess	***I. Vessels with dilatation***
1) discrete 2) patchy 3) diffuse	A) Roundish with dilatation B) Crowded or tortuous superficial vessels with dilatation
II. Erosions (size <5mm in diameter)	***II. Intramucosal bleeding***
1) discrete 2) patchy 3) diffuse	A) With round, dilated vessels B) With crowded or tortuous superficial vessels that are dilated
III. Ulceration (size >5mm in diameter)	***III. Intraluminal bleeding***
1) discrete 2) patchy 3) diffuse	A) With round, dilated vessels B) With crowded or tortuous superficial vessels that are dilated
B) PICaSSO-Ab. (*abbreviated version*)	
0. No mucosal defect	***0. Vessels without dilatation***
Any from: - Continuous / regular crypt architecture - Crypts not visible (scar) - Discontinuous and/or dilated/elongated crypts	Any from: - Roundish; following crypt architecture - Vessels not visible (scar) - Sparse (deep) vessels without dilatation
I. Micro-erosions or crypt abscess	***I. Vessels with dilatation***
- Any pattern of discrete/patchy/diffuse involvement:	- Vessels that are either roundish and dilated; or crowded and tortuous with dilatation
II. Erosions (size <5mm in diameter)	***II. Intramucosal bleeding***
III. Ulceration (size >5mm in diameter)	***III. Intraluminal bleeding***

Table 2: Intraclass correlation coefficients (ICC) pre- and post-training modules (Round 1)

	Pre-training ICC (95% CI)	**Post-training ICC (95% CI)**
Mayo score	0.775 (0.678 – 0.864)	0.818 (0.731 – 0.894)
UCEIS Total score	0.786 (0.692 – 0.872)	0.833 (0.753 – 0.903)
- UCEIS vascular pattern	0.429 (0.306 – 0.588)	0.417 (0.295 – 0.577)
- UCEIS bleeding pattern	0.689 (0.574 – 0.804)	0.726 (0.617 – 0.831)
- UCEIS erosion / ulcer pattern	0.770 (0.672 – 0.861)	0.810 (0.723 – 0.887)
PICaSSO mucosal component	0.754 (0.651 – 0.850)	0.848 (0.773 – 0.913)
PICaSSO-Ab. mucosal (abbreviated version)	0.754 (0.651 – 0.851)	0.826 (0.743 – 0.899)
PICaSSO vascular component	0.657 (0.536 – 0.781)	0.739 (0.631 – 0.842)
PICaSSO-Ab. vascular (abbreviated version)	0.622 (0.498 – 0.754)	0.746 (0.640 – 0.847)

ICCs are from a two-way random model with absolute agreement, and are for single measures (n = 15 participants). All values are significant at a $p<0.001$.

Abbreviations: CI, confidence interval; EVC, endoscopic virtual chromoendoscopy score; ICC, intraclass correlation coefficient; UCEIS, ulcerative colitis endoscopic index of disease severity

Table 3: Intraclass correlation coefficients (ICC) for the participants attending Round 1 and Round 2*

	Round 1 post-training ICC (95% CI)	Round 2 ICC (95% CI)
Mayo score	0.835 (0.745 – 0.906)	0.881 (0.823 – 0.926)
UCEIS Total score	0.844 (0.763 – 0.911)	0.881 (0.827 – 0.925)
- UCEIS vascular pattern	0.488 (0.349 – 0.650)	0.602 (0.487 – 0.722)
- UCEIS bleeding pattern	0.737 (0.623 – 0.841)	0.789 (0.704 – 0.863)
- UCEIS erosion / ulcer pattern	0.825 (0.739 – 0.898)	0.869 (0.812 – 0.917)
PICaSSO mucosal component	0.873 (0.805 – 0.929)	0.869 (0.812 – 0.917)
PICaSSO-Ab. mucosal (abbreviated version)	0.852 (0.774 – 0.916)	0.854 (0.792 – 0.907)
PICaSSO vascular component	0.715 (0.597 – 0.826)	0.783 (0.695 – 0.860)
PICaSSO-Ab. vascular (abbreviated version)	0.736 (0.622 – 0.840)	0.772 (0.682 – 0.851)

*Nine / 15 participants attended on both round 1 (30 videos) and round 2 (42 videos).

ICCs are from a two-way random model with absolute agreement, and are for single measures. All values are significant at a $p<0.001$.

Abbreviations: CI, confidence interval; EVC, endoscopic virtual chromoendoscopy score; ICC, intraclass correlation coefficient; UCEIS, ulcerative colitis endoscopic index of disease severity

Figure 1: Study overview

Flow chart of study design, illustrating construction and delivery of the training module, and subsequent assessment of reliability and performance accuracy. Histological strata are according to the New York Mt. Sinai System proposed by Hefti *et al.*[23]

Figure 2: Representative photos of PICaSSO mucosal and vascular endoscopic findings captured by electronic virtual chromoendoscopyin different grade of inflammation

Figure 3: Correlation between the EVC score and histology in ulcerative colitis

Graphical representation of score distribution for PICaSSO according to histological severity of disease. Results are shown for [A] the mucosal component and [B] vascular components of the EVC score. Each point represents the average score per video over the 15 participants (post-training). *N* numbers characterise the number of videos tested in each histological category in the post-training module. Histological strata are presented according to the New York Mount Sinai scoring system.[23]

Figure 4: Inter-rater reliability across histological strata

Inter-rater reliability for each score across individual histological strata measured on day one (post-training) are presented as percentage agreement for every possible combination of raters (15 participants; n =105 possible pairings).

* Data for abbreviated PICaSSO system (PICaSSO-Ab.) is shown

NYMSS, New York Mt. Sinai System; UCEIS, ulcerative colitis endoscopic index of severity

Figure 5: Discriminating between quiescent vs. mild histological disease

ROC curves indicating the diagnostic accuracy of each individual endoscopic scoring system in predicting quiescent histological disease activity vs. mild/moderate/severe disease collectively (total *n* = 450 individual rater responses; 15 participants and 30 videos in the post-training module) [**A];** and discriminated against mild histological disease only (total *n* = 255 individual rater responses; 15 participants and 17 videos in the post-training module).

The post-training module consisted of videos capturing quiescent (*n* = 10), mild (*n* = 7), moderate (*n* = 6) and severe (*n* = 7) histological disease. Single asterisk denotes statistical significance in ROC curve analysis with a p value of <0.001. AUROC, area under the ROC curve; CI, confidence interval; PICaSSO, Paddington International virtual ChromoendoScopy ScOre; ROC, receiver operator characteristic; UCEIS, ulcerative colitis index of disease severity.

Figure 6. Representative photos of different grade of inflammation captured using HD and NBI

Figure 7. Representative photo of different grade of inflammation captured using HD and BLI (Courtesy of A.Repici and F.Furfaro Humanitas University, Milan, Italy)

Figure 8. Representative photos of different grade of inflammation captured using HD and iSCAN OE

Supplementary Table 1: Percentage rater agreement across histological strata (post-training; round 1)

	Histological strata*				Overall agreement
	0	***1***	***2***	***3***	
Mayo score	56.1%	52.8%	51.9%	77.0%	59.4%
UCEIS total score	44.0%	31.2%	25.9%	35.5%	35.4%
- UCEIS vascular pattern	47.2%	48.8%	46.7%	58.1%	50.0%
- UCEIS bleeding pattern	93.9%	57.8%	36.3%	54.1%	64.7%
- UCEIS erosion/ulcer pattern	86.4%	76.7%	52.5%	52.8%	69.5%
PICaSSO mucosal (*full*)	39.6%	27.3%	22.5%	51.6%	36.1%
PICaSSO-Ab. mucosal (abbreviated version)	86.3%	79.2%	60.3%	83.1%	78.7%
PICaSSO vascular (*full*)	24.7%	22.6%	24.0%	49.1%	29.7%
PICaSSO-Ab. vascular (abbreviated version)	58.1%	42.7%	42.5%	67.3%	53.5%

Percentage (%) agreement across each histological strata determine by evaluating every possible rate combination (n = 15 raters; 105 possible rater combinations).
*Histological strata according to the New York Mt. Sinai scoring system.[23]

Supplementary Table 2A: Percentage rater agreement across histological strata (post-training; round 1)

	Histological strata*				Overall agreement
	0	***1***	***2***	***3***	
Mayo score	51.9%	47.6%	56.5%	86.5%	59.9%
UCEIS total score	38.6%	28.6%	25.9%	45.2%	35.3%
- UCEIS vascular pattern	42.8%	49.2%	48.6%	69.8%	51.8%
- UCEIS bleeding pattern	90.6%	54.8%	38.9%	60.7%	64.9%
- UCEIS erosion/ulcer pattern	85.6%	77.8%	52.3%	57.5%	70.6%
PICaSSO vascular; *full*	24.4%	24.6%	22.7%	52.0%	30.6%
PICaSSO vascular; *abbreviated*	54.2%	36.5%	43.1%	70.6%	51.7%
PICaSSO mucosal; *full*	43.3%	28.6%	24.5%	57.8%	39.4%
PICaSSO mucosal; *abbreviated*	82.2%	81.7%	62.0%	88.5%	79.5%

Percentage (%) agreement across each histological strata determine by evaluating every possible rate combination (*n = 9/15* raters who participated in both round 1 and round 2; 36 possible rater combinations).

*Histological strata according to the New York Mt. Sinai scoring system.[23]

Supplementary Table 2B: Percentage rater agreement across histological strata (round 2)

	Histological strata*				Overall agreement
	0	***1***	***2***	***3***	
Mayo score	70.0%	55.2%	53.2%	83.8%	69.0%
UCEIS total score	53.0%	33.0%	37.0%	34.6%	41.2%
- UCEIS vascular pattern	59.6%	59.4%	59.3%	73.9%	64.0%
- UCEIS bleeding pattern	94.4%	53.5%	64.4%	47.9%	67.9%
- UCEIS erosion/ulcer pattern	89.8%	61.1%	55.6%	63.0%	71.2%
PICaSSO vascular; *full*	29.6%	31.9%	20.4%	39.7%	31.9%
PICaSSO-Ab. vascular (abbreviated version)	67.0%	40.6%	45.8%	60.5%	56.9%
PICaSSO mucosal; *full*	42.0%	21.2%	17.1%	54.9%	38.5%
PICaSSO-Ab. mucosal (abbreviated version)	89.8%	63.2%	68.1%	82.1%	79.2%

Percentage (%) agreement across each histological strata determine by evaluating every possible rate combination (*n = 9/15* raters who participated in both round 1 and round 2; 36 possible rater combinations).

*Histological strata according to the New York Mt. Sinai scoring system.[23]

Supplementary Table 3: Accuracy of PICaSSO in discriminating quiescent vs. mild histological disease

Criterion	Cut-point	Sensitivity	Specificity
Mayo endoscopic sub score	0	68%	68%
UCEIS total score	3	99%	10%
- UCEIS vascular pattern	1	39%	68%
- UCEIS bleeding pattern	1	97%	32%
- UCEIS erosion/ulcer pattern	1	93%	47%
PICaSSO vascular	0C	83%	47%
PICaSSO mucosal	0C	97%	45%

PICaSSO, Paddington International virtual ChromoendoScopy ScOre; UCEIS, ulcerative colitis endoscopic index of disease severity.