Assessing the realism of colonoscopy simulation: the development of an instrument and systematic comparison of 4 simulators

Andrew Hill, LLB(Hons), BA(Hons),[1,4] Mark S. Horswill, PhD,[1] Annaliese M. Plooy, PhD,[3] Marcus O. Watson, PhD,[2,4] Rozemary Karamatic, MBBS,[2,4] Tabinda A. Basit, BA(Hons),[1] Guy M. Wallis, PhD,[3] Stephan Riek, PhD,[3] Robin Burgess-Limerick, PhD,[3] David G. Hewett, MBBS, MSc, PhD, FRACP[2,4]

Brisbane, Australia

Background: No useful comparative data exist on the relative realism of commercially available devices for simulating colonoscopy.

Objectives: To develop an instrument for quantifying realism and provide the first wide-ranging empiric comparison.

Design: Repeated measures, observational study. Nineteen experienced colonoscopists completed cases on 4 colonoscopy simulators (AccuTouch, GI Mentor II, Koken, and Kyoto Kagaku) and evaluated each device.

Setting: A medical simulation center in a large tertiary hospital.

Main Outcome Measures: For each device, colonoscopists completed the newly developed Colonoscopy Simulator Realism Questionnaire (CSRQ), which contains 58 items grouped into 10 subscales measuring the realism of different aspects of the simulation. Subscale scores are weighted and combined into an aggregated score, and there is also a single overall realism item.

Results: Overall, current colonoscopy simulators were rated as only moderately realistic compared with real human colonoscopy (mean aggregated score, 56.28/100; range, 48.39–60.45, where 0 = "extremely unrealistic" and 100 = "extremely realistic"). On both overall realism measures, the GI Mentor II was rated significantly less realistic than the AccuTouch, Kyoto Kagaku, and Koken ($P < .001$). There were also significant differences between simulators on 9 subscales, and the pattern of results varied between subscales.

Limitations: The study was limited to commercially available simulators, excluding ex-vivo models. The CSRQ does not assess simulated therapeutic procedures.

Conclusions: The CSRQ is a useful instrument for quantifying simulator realism. There is no clear "first choice" simulator among those assessed. Each has unique strengths and weaknesses, reflected in the differing results observed across 9 subscales. These findings may facilitate the targeted selection of simulators for various aspects of colonoscopy training. (Gastrointest Endosc 2012;75:631-40.)

Colonoscopy is complex and difficult to learn, and there is no clear consensus on how many colonoscopies are required during training for the attainment of competence.[1-3] Because colonoscopy insertion skill is taught primarily via the Halstedian apprenticeship model,[4,5] the attainment of basic competence incurs risks to patients' safety and comfort.[6,7] Hence, simulation training is potentially beneficial both in reducing the time it takes for trainees to achieve competence and in reducing the inherent risks to patients. The realism or fidelity of the simula-

Abbreviations: ANOVA, analysis of variance; CSRQ, Colonoscopy Simulator Realism Questionnaire; CSRQ-IS, Colonoscopy Simulator Realism Questionnaire—Importance Supplement; PM, physical model; VR, virtual reality.

DISCLOSURE: *Supported by the Australian Government Department of Health and Ageing and the National Health and Medical Research Council of Australia. Dr Hewett was supported by an NHMRC-NICS-Queensland Health fellowship, and Dr Wallis was supported by an Australian Research Council Future Fellowship (FT100100020). The funding sources had no role in the collection, analysis, or interpretation of data. All other authors disclosed no financial relationships relevant to this publication.*

0016-5107/$36.00
doi:10.1016/j.gie.2011.10.030

Received July 18, 2011; Accepted October 22, 2011.

Current affiliations: Schools of Psychology (1), Medicine (2), and Human Movement Studies (3), The University of Queensland, Clinical Skills Development Service, Queensland Health (4); Brisbane, Australia.

Reprint requests: David G. Hewett, MBBS, MSc, PhD, FRACP, The University of Queensland School of Medicine, Herston QLD, Australia 4006.

tion has been identified as an important prerequisite for a highly effective colonoscopy simulator.[8]

Commercially available devices for simulating colonoscopy fall into 3 categories: computer-based virtual reality (VR) simulators, physical models (PMs), and ex-vivo animal models. VR simulators, such as the Endoscopy AccuTouch System (Immersion Medical; Gaithersburg, MD) and the GI Mentor II (Simbionix Corp., Cleveland, OH), are typically stand-alone systems, equipped with their own monitors and modified or mock colonoscopes. By contrast, PM simulators, such as the Koken Colonoscopy Training Model Type 1-B (Koken Co. Ltd.; Tokyo, Japan) and the Kyoto Kagaku Colonoscope Training Model (Kyoto Kagaku Co. Ltd.; Kyoto, Japan), are designed for use in conjunction with genuine colonoscopy equipment. So too are ex-vivo animal models, such as the Endo X Trainer bovine colon model (Medical Innovations International Inc., Rochester, MN). In addition, ex-vivo animal models require a supply of fresh or frozen unperforated animal colons.[9]

Several studies have investigated the realism of individual colonoscopy simulators.[9-14] However, because these studies used different raters and a variety of measures and response modes (Likert scales, rating scales, or visual analogue scales), no useful comparative data exist on the relative realism of the available devices. What these studies have in common is that each involved experienced colonoscopists completing a series of response scales to quantify the realism of various aspects of a simulator, such as the visual representation of the mucosa or the haptic feedback.[9-13] However, despite the importance of deconstructing and assessing the realism of all key components of a simulator,[8] scant details were given of the processes by which the questions were derived, and the questionnaires used contained relatively few items (5-14 per questionnaire).

The aims of this study were to develop a comprehensive instrument for the measurement of colonoscopy simulator realism and to directly compare the realism of all commercially available VR and PM colonoscopy simulators.

METHODS

Our first step was to develop a new 59-item instrument for assessing the fidelity of colonoscopy simulators, the Colonoscopy Simulator Realism Questionnaire (CSRQ), and a complementary measure to assess the importance of each element of simulated colonoscopy covered by the CSRQ, the Colonoscopy Simulator Realism Questionnaire—Importance Supplement (CSRQ-IS). We then conducted an empiric study in which experienced colonoscopists performed a standardized repertoire of cases on each of 4 simulators before rating their realism using the CSRQ. The Human Research Ethics Committees of the Royal Brisbane and Women's Hospital and the University of Queensland approved the research.

Take-home Message

- The Colonoscopy Simulator Realism Questionnaire is a useful instrument for quantifying simulator realism, which may also aid the development of superior colonoscopy simulators. These results may facilitate the targeted selection of simulators for specific aspects of colonoscopy training.
- Improvements in colonoscopy simulators and simulation training may reduce both the time it takes trainees to achieve competence and the associated risks to patients' safety and comfort.

Colonoscopy simulators

We evaluated 4 colonoscopy simulators (Fig. 1), including 2 VR devices (the Endoscopy AccuTouch System and the GI Mentor II) and 2 PMs (the Koken Colonoscopy Training Model Type 1-B and the Kyoto Kagaku Colonoscope Training Model). All simulators were assembled according to the manufacturers' instructions. The PMs were lubricated and presented to the raters with their abdomen covers in place, atop a height-adjustable table (Fig. 2). Simulated PM colonoscopies used a high-definition endoscopy system (Exera II CLV-180 light source and CV-180 processor, OEV191H monitor, and CF-H180DL colonoscope; Olympus Medical Systems Corp., Tokyo, Japan), with air insufflation and suction. Table 1 provides a comparative overview of the 4 simulators. We chose not to evaluate an ex-vivo animal model in this study because the costs and logistics associated with sourcing, storing, and preparing animal colons are likely to preclude their widespread adoption for routine training purposes.

Colonoscopy Simulator Realism Questionnaire (CSRQ)

We developed the CSRQ to assess the realism of all key aspects of a colonoscopy simulator relevant to insertion and withdrawal, to provide a more comprehensive alternative to those used previously.[9-13] It was designed to apply to all 3 categories of simulators, and the online supplement (available at www.giejournal.org) describes its development.

The final CSRQ contains 59 items that require the rater to compare a colonoscopy simulator with real colonoscopy (58 assess the realism of specific aspects of the simulation, and 1 assesses overall realism). The responses are given on a rating scale ranging from 1 ("extremely unrealistic") to 6 ("extremely realistic"). To prevent ambiguous missing data, items that may not apply to all simulators also include a "not applicable" option. The 58 specific items can be divided into 10 a priori subscales relating to

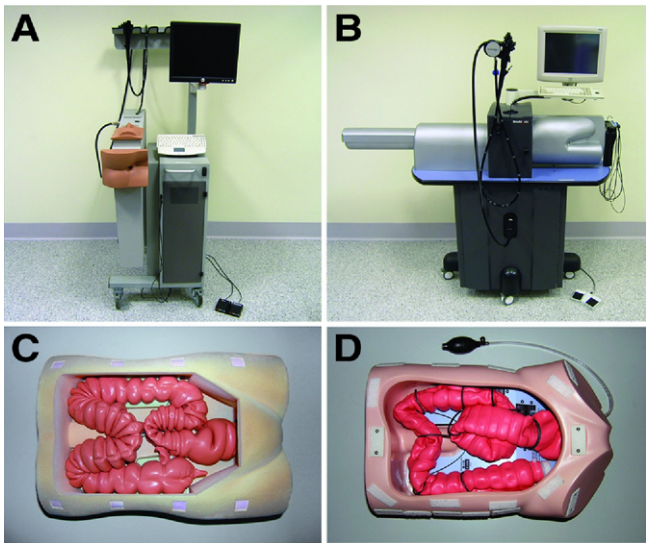

Figure 1. Virtual reality simulators and physical models assessed in this study. **A**, Immersion Endoscopy AccuTouch System. **B,** Simbionix GI Mentor II. **C,** Koken Colonoscopy Training Model Type 1-B (with abdomen cover removed). **B**, Kyoto Kagaku Colonoscope Training Model (with abdomen cover removed).

Figure 2. Laboratory setup used for the physical models, with the Kyoto Kagaku model in place.

different broad aspects of the simulation: (1) the *physical arrangement* of the equipment and the anus, (2) the *colonoscope*, (3) the *anatomic structure* of the colon, (4) the *visual representation* of the colon, (5) the *visual response* of the simulator, (6) the *haptic response* of the simulator, (7) *insufflation and deflation*, (8) the *difficulty of navigation*, (9) *looping*, and (10) *patient discomfort*. Supplementary Table 1 (available online at www.giejournal.org) presents all items, grouped by subscale, and the online supplement also includes a ready-to-use version of the questionnaire.

CSRQ Importance Supplement (CSRQ-IS)

The CSRQ-IS contains 58 items corresponding to the 58 specific CSRQ questions. Raters indicate how important they believe it is that a colonoscopy simulator used by trainees should perform well on each aspect of simulated colonoscopy (eg, "The realism of the *visual representation* of colonic mucosa"). Responses are given on a rating scale ranging from 1 ("extremely unimportant") to 6 ("extremely important"). A ready-to-use version is provided in the online supplement.

Raters

A convenience sample of 19 experienced colonoscopists participated as raters (gastroenterologists, n = 16; colorectal surgeons, n = 2; general surgeons, n = 1). On average, they had 12 years of experience in endoscopic practice (range, 3-33, SD = 8.39) and had completed 11,137 colonoscopies (range, 1000-50,000; SD = 13,732). Five had previous experience with 1 or more of the simulators (AccuTouch, n = 3; GI Mentor II, n = 3; Koken, n = 1; Kyoto Kagaku, n = 1). On explicit questioning, no participant disclosed a financial relationship with any company that manufactures or distributes colonoscopy training equipment. The recruitment and testing were conducted between March and September 2009.

Realism evaluation procedure

Each rater participated in a single session lasting approximately 2 hours. All sessions were conducted in a research laboratory at a medical simulation center, supervised by a research assistant. Each rater completed the CSRQ-IS and then performed simulated colonoscopy on the 4 simulators, completing the CSRQ for each simulator before progressing to the next. The order in which the simulators were presented to each rater was randomized with 2 constraints. Within each block of 4 consecutive raters, each simulator (1) appeared in each position (first, second, third, and fourth) once and (2) was preceded and followed only once by each other simulator. (Twenty colonoscopists originally participated; however, 1 did not complete the session and was excluded.)

Before each use of the simulator, we encouraged the rater (when possible) to customize the physical layout of the equipment to suit individual preferences (eg, adjusting the position or angle of the monitor, table height). For each case, the rater inserted the colonoscope to the cecum (or, for the Kyoto Kagaku cases, to the end of the mucosa) and then withdrew. We did not evaluate therapeutic procedures. Changes in the patient's position and the application of abdominal pressure were permitted for applicable simulators, and the research assistant aided when appropriate (she had been trained in the relevant procedures by a gastroenterologist, D.G.H.). However, all other

TABLE 1. Comparative overview of the features of the 4 colonoscopy simulators

	Virtual reality simulators		Physical models	
Feature	**AccuTouch**	**GI Mentor II**	**Koken**	**Kyoto Kagaku**
Stand-alone system (includes colonoscope and monitor)	✓	✓	X	X
Offers a choice of cases	✓	✓	X	✓
Colon has a cecum	✓	✓	✓	X
Simulates patient discomfort	✓	✓	X	X
Allows or simulates patient position changes	✓	X	X	✓
Allows or simulates application of manual pressure	✓	X	✓	✓
Magnetic imaging view available	X	✓	✓*	✓*
"External" view of colon available	✓	✓	✓†	✓†
Simulates lesions and other abnormalities	✓	✓	X	X
Simulates therapeutic procedures	✓	✓	X	X
Gives feedback on performance	✓	✓	X	X

*The use of a magnetic imaging view with either of the physical models requires additional equipment, namely, a magnetic colonoscope and a position detection system, such as ScopeGuide (Olympus Medical Systems Corp.; Tokyo, Japan).
†Both physical models have removable abdomen covers, allowing the exterior of the colon to be made visible during the procedure.

available aids (eg, magnetic imaging view, didactic advice) were prohibited.

The raters completed 2 cases on each simulator with the exception of the Koken model, which has a single fixed configuration and only 1 completed case. For the other simulators, the cases selected were AccuTouch "Introduction" cases 3 and 6, GI Mentor II "First Module" cases 2 and 7, and Kyoto Kagaku cases 3 and 6. These were chosen to satisfy 2 criteria: (1) each case involves looping, and (2) the average difficulty of cases is approximately equal between simulators. In addition, the AccuTouch and GI Mentor II cases were matched so that raters were exposed to a similar range of pathologic conditions in each simulator. After completing the CSRQ for a simulator, raters compared the overall difficulty of the case/s they had just completed with that of the average routine colonoscopy, using a rating scale ranging from 1 ("much easier") through 4 ("same") to 7 ("much more difficult").

Scoring of CSRQ ratings

The online supplement explains how we calculated the CSRQ subscale scores and the overall aggregated CSRQ scores, using CSRQ-IS responses to weight the ratings. Figure 3 presents the mean CSRQ-IS importance score for each subscale.

Statistical analyses

We compared the simulators on 3 outcome measures: CSRQ subscale scores, aggregated CSRQ scores, and responses to the "overall realism" item. We used 1-way repeated measures analyses of variance (ANOVAs) for these procedures and also to analyze responses to the "overall difficulty" control item. For these omnibus tests, alpha was set at .05, and Mauchly's test was used to identify violations of the sphericity assumption. For tests in which this assumption was breached, we used the conservative Greenhouse-Geisser correction to adjust the degrees of freedom.[15] For each statistically significant omnibus test result, η^2 was calculated to quantify the effect size[15] (η^2 indicates the proportion of the variation in ratings explained by differences between the simulators in our sample). Each significant ANOVA was followed up with 6 pairwise comparisons (comparing each simulator with every other), and we applied the Bonferroni-Holm correction for multiple comparisons[16] to maintain the family-wise error rate at $P < .05$. Additional analyses verified that the study outcomes were not affected by interrater differences in experience with colonoscopy or with the simulators that were evaluated (see the online supplement for details). We used SPSS 16.0 (SPSS Inc.; Chicago, IL) for all analyses.

We assessed the interrater reliability for each outcome measure by transposing the data so that scores for the 4 simulators were arranged in rows with 1 column per rater, and then calculating Cronbach's alpha across all raters.[17] It was inappropriate to factor analyze the CSRQ or to assess the reliability of the instrument or its subscales through measures of internal consistency for 2 reasons. First, unlike with traditional questionnaire instruments, the simulators were the units of analysis rather than the respondents. Second, because the CSRQ was designed to be compre-

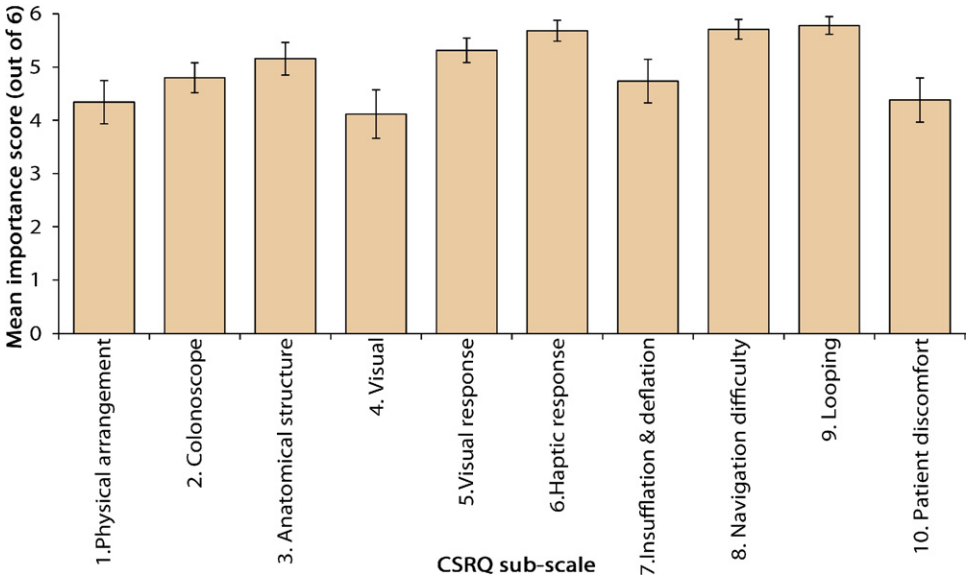

Figure 3. Mean importance scores out of 6 for each Colonoscopy Simulator Realism Questionnaire (CSRQ) realism subscale (with 95% confidence intervals).

hensive, there is no a priori reason to assume that a particular simulator's performance (relative to other simulators) should be consistent across all the items in a subscale. For example, in terms of visual realism, a simulator with a realistic-looking sigmoid colon does not necessarily have an equally realistic-looking cecum.

RESULTS

Overall realism

In comparison with real colonoscopy, the realism of all 4 colonoscopy simulators was rated as relatively low, with a mean aggregated CSRQ score of 56.28/100 (range, 48.39-60.45) (Fig. 4). There was a significant difference among the simulators for both measures of overall realism: aggregated CSRQ scores, $F(3,54) = 7.62$, $P < .001$, $\eta^2 = .30$; and ratings on the single "overall realism" item, $F(3,54) = 9.53$, $P < .0001$, $\eta^2 = .35$. Figure 4 indicates that both measures yielded the same pattern of results, which was confirmed by the pairwise comparisons. In each case, the GI Mentor II was rated significantly less realistic than the AccuTouch, the Kyoto Kagaku, and the Koken, and there was no significant difference in ratings among the latter 3 simulators. Interrater reliability was .87 for the aggregated CSRQ scores and .90 for the overall realism item.

CSRQ subscales

Figure 5 shows the mean CSRQ subscale scores for each simulator. Table 2 provides additional detail, including interrater reliabilities, ANOVA results, effect sizes, and a summary of pairwise comparisons. There were significant differences between simulators on 9 of the 10 subscales (the exception being navigation difficulty), and the pattern of results varied from subscale to subscale (Table 2).

Difficulty control item

Consistent with our aim to select cases of approximately equivalent difficulty, a 1-way repeated-measures ANOVA revealed no statistically significant difference in difficulty ratings among the 4 simulators, $F(3,54) = 2.71$, $P > .05$.

DISCUSSION

To our knowledge, this study is the first to provide a wide-ranging empiric comparison of the realism of several colonoscopy simulators as rated by experienced colonoscopists. Two separate measures of overall realism (a single item and a weighted aggregate score) yielded the same pattern of results: the GI Mentor II was rated as significantly less realistic than the AccuTouch, Kyoto Kagaku, and Koken simulators, which received similar ratings.

This finding provides support for the content validity of the CSRQ. Both measures produced the same pattern of results, which suggests that, when weighted and combined appropriately, the 58 CSRQ items that assess specific aspects of simulator realism were sufficiently comprehensive to provide a valid means of quantifying the overall realism of these 4 simulators. (However, this does not imply that the single overall realism item is a sufficient alternative to the full CSRQ, given that the process of responding to the more specific items may have prompted raters to consider aspects of realism they would have otherwise overlooked when generating

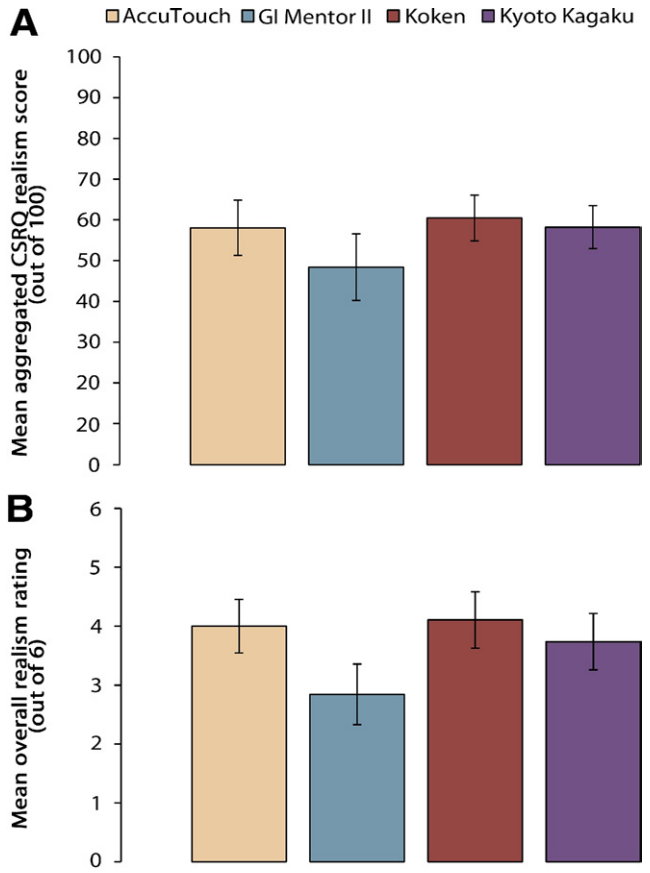

Figure 4. Two broad measures of simulator realism. **A**, Aggregated Colonoscopy Simulator Realism Questionnaire (CSRQ) realism scores. **B**, Ratings on a single item assessing overall realism (both with 95% confidence intervals).

overall ratings.) Further, because reliability is an important prerequisite to questionnaire validity, it is notable that the interrater reliability estimates for the CSRQ and 9 of its subscales were very high.

Our findings regarding overall realism are also relevant to the validity of the simulators. For example, they suggest that validation data for the GI Mentor II demonstrating a relationship between experience and performance[11,18-20] should not be interpreted in isolation. Although this relationship implies that performance on simulated and real colonoscopies share common skill elements, these studies do not demonstrate that the GI Mentor II faithfully replicates the task or that efforts to develop higher-fidelity simulators are without value. Indeed, others have argued that, contrary to commercial imperatives, the realism of an endoscopy simulator should be refined fully before validation efforts even begin.[8] Given the other relatively low aggregated CSRQ scores, similar caveats apply to efforts to validate the other 3 simulators. At present, such research has been published for the AccuTouch[10,21] but not the Kyoto Kagaku or the Koken.

The CSRQ subscales tell more of the story. Unsurprisingly, given the heterogeneity of the 4 simulators, different patterns of results were observed for different subscales. We now examine the results for each subscale to determine the relative strengths and weaknesses of each device.

For the *physical arrangement* subscale, the VR simulators were significantly outperformed by the PM systems. Because the PMs are designed for use with standard colonoscopy equipment, they can be arranged to closely resemble the layout of a genuine procedure room. By contrast, each VR simulator is packaged as a complete system that can easily be wheeled from one training room to another but the arrangement of its components (monitor, anus, colonoscope, etc.) is more constrained than in real life. The realism may be improved by partially disassembling the system and placing the monitor in a more orthodox and ergonomically appropriate position.[21]

For realism of the *colonoscope*, the PM systems again significantly outperformed the VR simulators. The ratings for both PMs were close to ceiling, which is unsurprising, given that they were used in conjunction with a genuine unmodified colonoscope. However, both VR simulators performed reasonably well and were rated higher on this subscale than on any other.

For the *anatomic structure* subscale, the results were different because the simulators were not split along VR versus PM lines. Among PMs, the Koken outperformed the Kyoto Kagaku. However, the relative realism of the Koken's anatomic structure comes at a cost: a single fixed configuration that offers no variability of practice. Among the VR simulators, the AccuTouch was rated as more realistic than the GI Mentor II. The trends in responses to the individual questionnaire items suggest that the degree of angulation at major junctions and flexures was considered less realistic in the GI Mentor II than in the AccuTouch.

In terms of *visual realism*, the raters assessed both VR simulators more favorably than the PMs. However, all 4 simulators were rated quite low on this subscale. The Kyoto Kagaku, which performed significantly worse than the other 3 simulators, does not even attempt to simulate several of the features assessed for visual realism (including the appearance of vascular patterns, the cecum, the appendiceal orifice, and the ileocecal valve), and, like the Koken, its mucosa has an unrealistically uniform, plastic appearance.

Despite this, the PMs generally outperformed the VR simulators in terms of the realism of their *visual response* to colonoscope advancement and steering maneuvers. For the VR systems, performance on this subscale is largely a matter of software engineering, whereas for the PMs it depends primarily on the physical properties of the colon (including lubrication), given that the colonoscope itself is genuine.

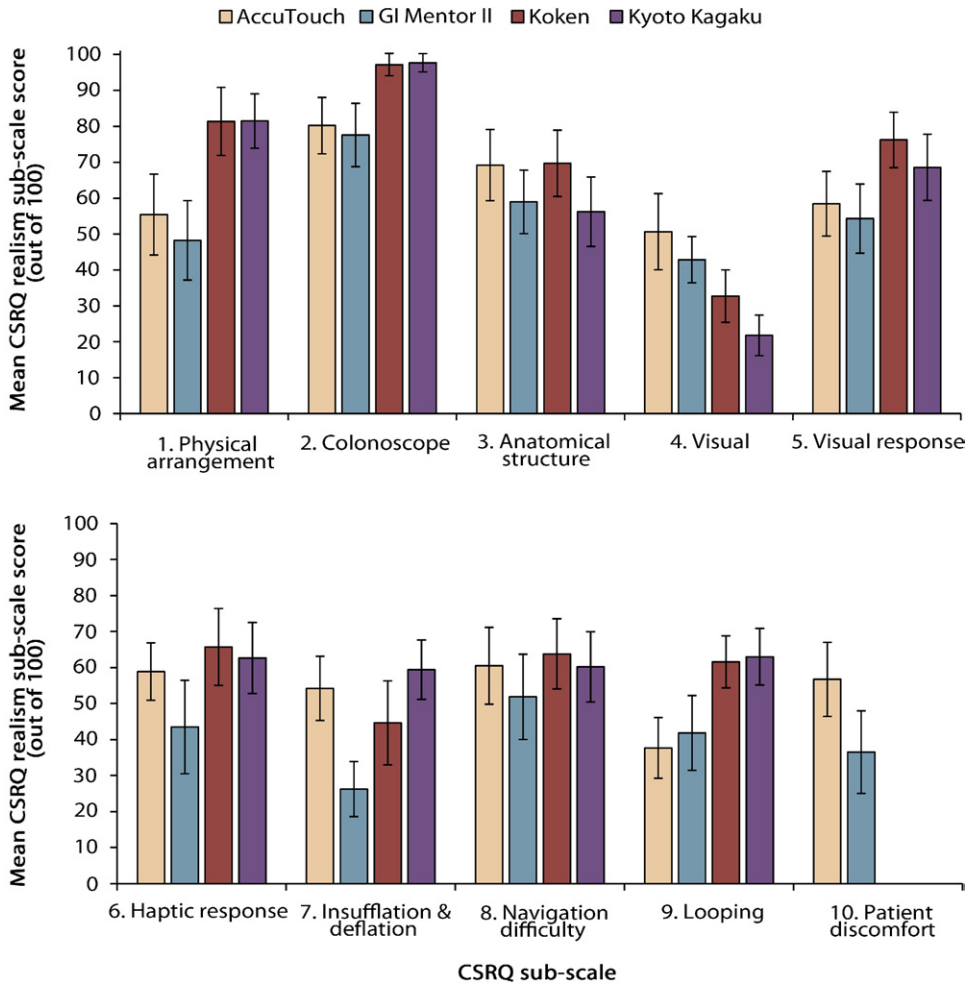

Figure 5. Mean Colonoscopy Simulator Realism Questionnaire (CSRQ) realism subscale scores out of 100 for each colonoscopy simulator (with 95% confidence intervals).

Similarly, the PMs have an inherent advantage in terms of simulating *haptic response*. Indeed, both PMs significantly outperformed the GI Mentor II, as did the AccuTouch. Both VR simulators have previously been criticized for the quality of their haptic feedback.[11,22-25] At the individual-item level, the GI Mentor II consistently received the lowest average rating of all 4 simulators, and it performed particularly poorly in relation to the amount of forward insertion force required and the sensation of resistance to movement of the colonoscope shaft. In a VR simulator, haptic response depends heavily on both software engineering and the quality and design of the haptic device and colonoscope.[24] Further refinement of these elements will be crucial if future VR simulators are to provide authentic force feedback to trainees.[22-24]

For simulation of *insufflation and deflation*, the GI Mentor II was again rated significantly less realistic than the other 3 systems, receiving a lower score than on any other subscale. The GI Mentor II received the lowest average rating of the 4 simulators on every individual item in this subscale (Supplementary Table 1). Among the others, none scored particularly well, and the Kyoto Kagaku, which was the highest rated, significantly outperformed the Koken, whereas the AccuTouch did not differ significantly from either PM.

For the *navigation difficulty* subscale, there was no significant difference between simulators. This may reflect our deliberate choice of cases of equivalent difficulty across all simulators in our aim to eliminate any potential confound between case difficulty and perceived realism. Further, navigation difficulty had one of the highest mean importance scores of all subscales. We retained this CSRQ subscale because it may prove informative to future researchers—for example, in a pre–post evaluation of 2 incremental versions of a simulator under development.

By contrast, the subscale assessing the realism of *looping* not only received the highest mean importance score of all 10 subscales but also yielded a significant difference between simulators. Both VR simulators scored poorly and were rated significantly lower than

TABLE 2. Mean CSRQ realism subscale scores out of 100 for each colonoscopy simulator, with significance tests for differences in ratings between simulators, effect sizes (η^2), and interrater reliabilities

CSRQ subscale	Simulator: mean CSRQ realism subscale score out of 100 (*SD*)				ANOVA omnibus *F*-test; summary of pairwise comparisons (Bonferroni-Holm correction)*	η^2	Interrater reliability
	AT	GIM	Ko	KK			
Physical arrangement	55.44 (22.36)	48.23 (22.94)	81.37 (19.67)	81.48 (15.69)	$F(3,54) = 27.12$, $P < .0001$; (GIM = AT) < (Ko = KK)	.60	.96
Colonoscope	80.20 (16.29)	77.59 (18.28)	97.19 (6.44)	97.69 (5.28)	$F(1.77,31.88) = 19.14$, $P < .0001$;† (GIM = AT) < (Ko = KK)	.52	.95
Anatomic structure	69.19 (19.18)	58.97 (20.08)	69.69 (18.34)	56.22 (20.57)	$F(3,54) = 5.39$, $P < .01$; KK < Ko, GIM < AT	.23	.81
Visual	50.64 (22.01)	42.86 (13.36)	32.69 (15.16)	21.77 (11.70)	$F(3,54) = 25.38$, $P < .0001$; KK < Ko < (GIM = AT)	.59	.96
Visual response	58.43 (19.73)	54.28 (20.00)	76.21 (15.99)	68.57 (19.11)	$F(3,54) = 9.11$, $P < .0001$; (GIM = AT) < Ko, GIM < KK	.34	.89
Haptic response	58.88 (16.51)	43.47 (26.92)	65.70 (22.14)	62.66 (20.45)	$F(2.26,40.62) = 7.55$, $P < .01$;† GIM < (AT = KK = Ko)	.30	.87
Insufflation & deflation	54.21 (18.44)	26.26 (15.87)	44.63 (24.23)	59.39 (17.08)	$F(3,54) = 16.08$, $P < .0001$; GIM < (Ko = AT, AT = KK), Ko < KK	.47	.94
Navigation difficulty	60.48 (22.12)	51.87 (24.55)	63.78 (20.23)	60.19 (20.23)	$F(3,54) = 2.32$, $P > .05$; N/A (GIM = KK = AT = Ko)	N/A	.57
Looping	37.67 (17.53)	41.85 (21.51)	61.58 (15.02)	63.00 (16.26)	$F(1.93,34.67) = 19.12$, $P < .0001$;‡ (AT = GIM) < (Ko = KK)	.52	.95
Patient discomfort	56.72 (21.32)	36.52 (23.80)	0 (0)	0 (0)	$F(3,54) = 70.01$, $P < .0001$; (KK = Ko) < GIM < AT	.80	.99

CSRQ, Colonoscopy Simulator Realism Questionnaire; *ANOVA,* analysis of variance; *AT,* AccuTouch; *GIM,* GI Mentor II; *Ko,* Koken; *KK,* Kyoto Kagaku.
*These summaries illustrate which mean differences between simulators were statistically significant by using the Bonferroni-Holm correction to maintain the family-wise error rate at $P < .05$. For example, "(KK = Ko) < GIM < AT" indicates that there was no significant difference between the Kyoto Kagaku and Koken models, that the GI Mentor II significantly outperformed both, and that the AccuTouch significantly outperformed all 3.
†For these analyses, the Greenhouse-Geisser correction was applied to the degrees of freedom to control for a violation of the sphericity assumption, as detected via Mauchly's test.

the PMs. These findings are consistent with previous observations.[5] The mean importance score also emphasizes the significance of looping for the fidelity of simulated colonoscopy.[5,26] Given the immense challenges involved in realistically simulating looping in a VR platform, PM simulators may offer a simple, cost-effective, and relatively realistic representation of this fundamental element of the overall task.

The final subscale assessed the realism of *patient discomfort.* Because neither PM attempts to simulate this aspect of colonoscopy, both were outperformed by the VR simulators. Among these, the AccuTouch, despite scoring poorly, was rated more favorably than the GI Mentor II. Trends in response to individual questionnaire items suggest that both devices provided poor visual feedback but that the AccuTouch had more realistic auditory feedback and feedback on the *amount* of simulated patient discomfort.

Overall, our results suggest that there is no clear "first choice" colonoscopy simulator among the AccuTouch, the GI Mentor II, the Koken, and the Kyoto Kagaku models. The costs of colonoscopy simulators, particularly the VR systems, are not insubstantial. Therefore, training program directors, when faced with a choice among these 4 devices, should consider which subscales and which simulator features (Table 1) are most relevant to their specific educational aims. For example, our results suggest that PMs would be preferable to VR simulators for training on loop reduction techniques. However, if trainees need to be exposed to a diverse range of cases, then the Koken is not ideal. In fact, although the other 3 simulators do offer a selection of cases, none offers anything close to the number of cases that a trainee is likely to require to achieve competence.[2,3,27,28] This may explain why the impact of the AccuTouch[29-31] or the GI Mentor II[32] on colonoscope insertion skills in trainees is relatively short-lived; superior performance during real colonoscopy (relative to peers who have not received simulation training) appears largely restricted to the early part of the learning curve.[7,29,32-34] Given the demonstrated value of varied practice in the acquisition of motor skills,[35] it is likely that future simulators offering a broader range of more realistic cases will yield greater advances toward competence in new trainees before live cases need even be attempted.

Our study has demonstrated the utility of the Colonoscopy Simulator Realism Questionnaire. For the 2 measures of overall realism and the 9 subscales that yielded significant differences between simulators, the effect sizes (as measured by η^2) were all large,[36] indicating the sensitivity of the instrument, which complements its demonstrated reliability and validity.

The primary limitation of our study is that the CSRQ does not assess the realism of simulated therapeutic procedures. Indeed, the importance and complexity of these subtasks warrants the future development of comprehensive instruments to specifically assess the realism of individual procedures, such as polypectomy. Another limitation is we did not include simulators that are not commercially available.[12-14,22-24,37] Likewise, although we deliberately designed the CSRQ to be equally appropriate for evaluating ex-vivo models, we did not include them in this study.

In future, others may wish to use the CSRQ to gather comparable data on other existing simulators, such as the Endo X Trainer bovine colon model,[9] or on new devices. We recommend also including in these studies at least 1 of the simulators that we examined, to provide a benchmark for comparison. Another potential use for the CSRQ is in the pre–post evaluation of incremental prototype versions of new colonoscopy simulators under development. Although a more rudimentary instrument has been previously used for this purpose,[13] feedback from the CSRQ would provide more potentially informative detail, especially if the results are scrutinized at the individual-item level. For this purpose, we recommend that raters be kept blind to any simulator enhancements and that they be randomly and blindly assigned to 2 conditions after the test: half rating the new simulator, half rerating the older version. If the CSRQ is used according to these recommendations, future efforts to improve the realism of colonoscopy simulators will be more sharply focused and likely to be more successful.

REFERENCES

1. Cass OW, Freeman ML, Peine CJ, et al. Objective evaluation of endoscopy skills during training. Ann Intern Med 1993;118:40-4.
2. Vargo JJ. North of 100 and south of 500: where does the "sweet spot" of colonoscopic competence lie? Gastrointest Endosc 2010;71:325-6.
3. Spier BJ, Benson M, Pfau PR, et al. Colonoscopy training in gastroenterology fellowships: determining competence. Gastrointest Endosc 2009;71:319-24.
4. Reznick RK, MacRae H. Teaching surgical skills: changes in the wind. N Engl J Med 2006;355:2664-9.
5. Williams CB, Thomas-Gibson S. Rational colonoscopy, realistic simulation, and accelerated teaching. Gastrointest Endosc Clin N Am 2006;16:457-70.
6. Sedlack RE. Endoscopic simulation: where we have been and where we are going. Gastrointest Endosc 2005;61:216-8.
7. Sedlack RE. Simulators in training: defining the optimal role for various simulation models in the training environment. Gastrointest Endosc Clin N Am 2006;16:553-63.
8. Cohen J, Nuckolls L, Mourant RR. Endoscopy simulators: lessons from the aviation and automobile industries. Gastrointest Endosc Clin N Am 2006;16:407-23.
9. Sedlack RE, Baron TH, Downing SM, et al. Validation of a colonoscopy simulation model for skills assessment. Am J Gastroenterol 2007;102:64-74.
10. Sedlack RE, Kolars JC. Validation of a computer-based colonoscopy simulator. Gastrointest Endosc 2003;57:214-8.
11. Koch AD, Buzink SN, Heemskerk J, et al. Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. Surg Endosc 2008;22:158-62.
12. Koch AD, Haringsma J, Schoon EJ, et al. A second-generation virtual reality simulator for colonoscopy: validation and initial experience. Endoscopy 2008;40:735-8.
13. Yi SY, Woo HS, Ahn W, et al. Clinical evaluation of the KAIST-Ewha Colonoscopy Simulator II. Stud Health Technol Inform 2007;125:512-4.
14. Haycock AV, Bassett P, Bladen J, et al. Validation of the second-generation Olympus colonoscopy simulator for skills assessment. Endoscopy 2009;41:952-8.
15. Howell DC. Statistical methods for psychology. 4th ed. Belmont, CA: Duxbury Press, 1997.
16. Holm S. A simple sequential rejective multiple test procedure. Scand J Stat 1979;6:65-70.
17. MacLennan RN. Interrater reliability with SPSS for Windows 5.0. Am Stat 1993;47:292-6.
18. Ritter EM, McClusky DA 3rd, Lederman AB, et al. Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. J Gastrointest Surg 2003;7:871-7.
19. Felsher JJ, Olesevich M, Farres H, et al. Validation of a flexible endoscopy simulator. Am J Surg 2005;189:497-500.
20. Grantcharov TP, Carstensen L, Schulze S. Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. JSLS 2005;9:130-3.
21. Moorthy K, Munz Y, Orchard TR, et al. An innovative method for the assessment of skills in lower gastrointestinal endoscopy. Surg Endosc 2004;18:1613-9.
22. Woo HS, Kim WS, Ahn W, et al. Improved haptic interface for colonoscopy simulation. Conf Proc IEEE Eng Med Biol Soc 2007:1253-6.
23. Woo HS, Kim WS, Ahn W, et al. Haptic interface of the KAIST-Ewha colonoscopy simulator II. IEEE Trans Inf Technol Biomed 2008;12:746-53.
24. Maillard P, Flaction L, Samur E, et al. Instrumentation of a clinical colonoscope for surgical simulation. Conf Proc IEEE Eng Med Biol 2008:70-3.
25. Dogramadzi S, Virk GS, Bell GD, et al. Recording forces exerted on the bowel wall during colonoscopy: in vitro evaluation. Int J Med Robot 2005;1:89-97.
26. Thomas-Gibson S, Williams CB. Colonoscopy training: new approaches, old problems. Gastrointest Endosc Clin N Am 2005;15:813-27.
27. Church J, Oakley J, Milsom J, et al. Colonoscopy training: the need for patience (patients). ANZ J Surg 2002;72:89-91.
28. Tassios PS, Ladas SD, Grammenos I, et al. Acquisition of competence in colonoscopy: the learning curve of trainees. Endoscopy 1999;31:702-6.
29. Sedlack RE, Kolars JC. Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. Am J Gastroenterol 2004;99:33-7.
30. Park J, MacRae H, Musselman LJ, et al. Randomized controlled trial of virtual reality simulator training: transfer to live patients. Am J Surg 2007;194:205-11.
31. Ahlberg G, Hultcrantz R, Jaramillo E, et al. Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? Endoscopy 2005;37:1198-204.
32. Cohen J, Cohen SA, Vora KC, et al. Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. Gastrointest Endosc 2006;64:361-8.

33. Sturm LP, Windsor JA, Cosman PH, et al. A systematic review of skills transfer after surgical simulation training. Ann Surg 2008;248:166-79.
34. Tsuda S, Scott D, Doyle J, et al. Surgical skills training and simulation. Curr Probl Surg 2009;46:271-370.
35. Schmidt R, Lee TD. Motor control and learning: a behavioral emphasis. 5th ed. Champaign, IL: Human Kinetics, 2011.
36. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L. Erlbaum Associates, 1988.
37. Haycock A, Koch AD, Familiari P, et al. Training and transfer of colonoscopy skills: a multinational, randomized, blinded, controlled trial of simulator versus bedside training. Gastrointest Endosc 2010; 71:298-307.

SUPPLEMENTARY ONLINE CONTENT

Development of the Colonoscopy Simulator Realism Questionnaire (CSRQ)

After reviewing existing instruments[1-5] as a starting point, we examined and trialed the four simulators to deconstruct their key components and identify aspects of simulated colonoscopy not specifically addressed in previous measures. Next, we created an initial 38-item draft questionnaire and interviewed 8 experienced colonoscopists, who were attending a national gastroenterology scientific meeting, to elicit feedback on the draft. The interviewees reported a mean experience of 11,187 colonoscopies (range 2,000 to 25,000; $SD = 7{,}819$) and mean prior exposure to 3 different colonoscopy simulators (range 1 to 5; $SD = 1.51$). Colonoscopists provided detailed feedback on the wording of questions, and were asked to nominate additional aspects of realism that were not covered by the draft questionnaire. A comprehensive list of their comments was generated from interview transcripts, with each unique relevant suggestion used to generate a new item, bringing the total to 63. Of these, four items were dropped from the final questionnaire after data were collected (one was rated much less important than all other items, and three yielded large quantities of missing data because they related to functionality not used by a majority of the raters during the study).

Scoring of CSRQ ratings

To generate *CSRQ sub-scale scores*, we completed the following steps separately for each rater for each simulator. First, any "not applicable" responses to CSRQ items were scored as 1 (i.e., the lowest score on the scale). Second, *weighted realism scores* out of 36 for each CSRQ item were calculated by multiplying each CSRQ item rating by its corresponding CSRQ-IS importance rating. Third, we derived *preliminary sub-scale scores* by averaging the weighted realism scores for the component items. Finally, to ensure comparability between raters, we re-scaled the preliminary subscale scores into a final *CSRQ sub-scale score* (range 0 to 100), according to the following formula:

$$subscale\ score = \frac{(preliminary\ subscale\ score) - (mean\ importance\ rating \times 1)}{(mean\ importance\ rating \times 6) - (mean\ importance\ rating \times 1)} \times 100$$

In the above formula, "mean importance rating × 1" represents the rater's minimum possible preliminary subscale score, given the importance ratings that they had nominated. Similarly, "mean importance rating × 6" is equivalent to the rater's maximum possible preliminary subscale score.

We derived *aggregated CSRQ scores* for each simulator using a two-step process. First, we calculated the mean importance score out of 6 for each CSRQ sub-scale across our final sample of 19 colonoscopists (see Supplementary Table 1 for item means and Figure 3 for subscale means). Second, we used these scores as *sub-scale weightings* to calculate a final *aggregated CSRQ score* out of 100 for each rater according to the following formula:

$$aggregated\ score = \frac{\Sigma(subscale\ score \times subscale\ weighting)}{\Sigma(subscale\ weighting)}$$

In contrast to simply averaging responses to all CSRQ items, this method prevented the more populous subscales from contributing disproportionately to the aggregated scores, instead ensuring that the contribution of each sub-scale was proportional to its importance.

Statistical Checks on Experience Effects

To verify that the study outcomes were not affected by inter-rater differences in experience with colonoscopy, or with the simulators that were evaluated, we repeated the ANOVAs described in the Statistical Analyses section for each outcome measure, with two small changes. In one version of these analyses, the number of past colonoscopies performed by each rater was added as a covariate. In the other version, we compared raters who had previously used one or more of the simulators with those who had not by adding a new between-participants factor to each ANOVA.

When the ANOVAs for each outcome measure were re-run with the number of past colonoscopies added as a covariate, there was no substantive change to the existing effects, and colonoscopic experience had no main or interactive effect for any outcome measure (all p's $> .05$). Likewise, when simulator experience was added as a factor in the analyses, it had no main or interactive effects (all p's $> .05$). In these analyses, the substantive results appeared to change for one measure: The main effect of *simulator* on ratings of *navigation difficulty* became statistically significant, $F(3,51) = 3.62$, $p = .02$. However, all pairwise comparisons were non-significant after correction, so we dismissed this change as a chance occurrence. Hence, there was no evidence that any of the study outcomes were affected by experience differences between raters.

REFERENCES

1. Sedlack RE, Kolars JC. Validation of a computer-based colonoscopy simulator. Gastrointest Endosc 2003;57:214-8.
2. Sedlack RE, Baron TH, Downing SM, Schwartz AJ. Validation of a colonoscopy simulation model for skills assessment. Am J Gastroenterol 2007; 102:64-74.
3. Koch AD, Buzink SN, Heemskerk J, Botden SM, Veenendaal R, Jakimowicz JJ, Schoon EJ. Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. Surg Endosc 2008;22:158-62.
4. Koch AD, Haringsma J, Schoon EJ, de Man RA, Kuipers EJ. A second-generation virtual reality simulator for colonoscopy: validation and initial experience. Endoscopy 2008;40:735-8.
5. Yi SY, Woo HS, Ahn W, Kim WS, Lee DY. Clinical evaluation of the KAIST-Ewha Colonoscopy Simulator II. Stud Health Technol Inform 2007;125:512-4.

Supplementary Table 1. CSRQ items grouped by realism sub-scale, with mean importance ratings

***Realism Sub-scale* Items**	**Mean importance, 1 to 6 (SD)**
1. Physical Arrangement	
1. How realistic was the spatial arrangement of the monitor, colonoscope, and anus (after any adjustments you may have made)?	4.74 (0.87)
2. How realistic was the position of the monitor (after any adjustments you may have made)?	4.32 (0.89)
3. How realistic was the position of the anus (after any adjustments you may have made)?	4.47 (1.12)
4. How realistic was the position of the socket that the colonoscope plugs into (after any adjustments you may have made)?	3.84 (1.17)
2. Colonoscope	
5. How realistic was the weight of the colonoscope control head?	4.95 (0.71)
6. How realistic was the appearance of the colonoscope control head?	5.00 (1.00)
7. How realistic was the weight of the colonoscope insertion tube?	4.95 (0.91)
8. How realistic was the appearance of the colonoscope insertion tube?	4.05 (1.18)
9. How realistic was the length of the colonoscope insertion tube?	5.05 (0.91)
3. Anatomical Structure	
10. How realistic was the length of the colon?	5.00 (1.05)
11. How realistic was the degree of angulation at the rectosigmoid junction?	5.21 (0.54)
12. How realistic was the degree of angulation at the sigmoid-descending colon junction?	5.26 (0.56)
13. How realistic was the degree of angulation at the splenic flexure?	5.16 (0.76)
14. How realistic was the degree of angulation at the hepatic flexure?	5.16 (0.69)
4. Visual	
15. How realistic was the visual representation of colonic mucosa?	3.95 (0.97)
16. How realistic was the visual representation of vascular patterns?	3.58 (1.46)
17. How realistic was the visual representation of lesions and abnormalities?	4.21 (1.23)
18. How realistic was the visual representation of the light reflex?	4.21 (1.18)
19. How realistic was the visual representation of fecal residue?	3.47 (1.07)
20. How realistic was the visual representation of haustra/folds?	4.16 (1.30)
21. How realistic was the appearance of the rectum?	3.84 (1.21)
22. How realistic was the appearance of the sigmoid colon?	3.84 (1.21)
23. How realistic was the appearance of the descending colon?	3.79 (1.23)
24. How realistic was the appearance of the transverse colon?	4.00 (1.15)
25. How realistic was the appearance of the ascending colon?	3.79 (1.13)
26. How realistic was the appearance of the cecum?	4.89 (0.88)
27. How realistic was the appearance of the appendiceal orifice?	4.53 (0.96)
28. How realistic was the appearance of the ileocecal valve?	4.84 (0.96)
29. How realistic was the appearance of the terminal ileum?	4.32 (1.06)
30. Overall, how realistic was the visual representation of the colon?	4.47 (1.17)

Supplementary Table 1. Continued

***Realism Sub-scale* Items**	**Mean importance, 1 to 6 (SD)**
5. Visual Response	
31. How realistic was the response of the visual image when you advanced the scope?	5.42 (0.51)
32. How realistic was the response of the visual image to steering maneuvers?	5.47 (0.61)
47. How realistic was the response of mucosal folds to subtle steering maneuvers of the colonoscope?	5.05 (0.91)
6. Haptic Response	
33. How realistic was the amount of forward insertion force required?	5.74 (0.45)
34. How realistic was the amount of "torque" (clockwise or counter-clockwise rotational force) required?	5.79 (0.42)
35. How realistic was the feel of resistance to movement of the colonoscope shaft?	5.79 (0.42)
36. How realistic was the feel of resistance to movement of the colonoscope steering controls?	5.42 (0.77)
37. Overall, how realistic was the feel of resistance to movement of the colonoscope?	5.68 (0.48)
7. Insufflation and Deflation	
38. How realistic was the visual representation of air insufflation?	4.84 (0.60)
39. How realistic was the visual representation of air deflation?	4.89 (0.57)
40. How realistic was the feel of the air/water button?	4.37 (1.34)
41. How realistic was the visual representation when suction was applied?	5.00 (1.20)
42. How realistic was the feel of the suction button?	4.47 (1.31)
45. How realistic was the response of mucosal folds to air insufflation?	4.74 (1.10)
46. How realistic was the response of mucosal folds to suction?	4.84 (1.12)
8. Navigation Difficulty	
43. How realistic was the difficulty of navigating the colonoscope around bends and angulations?	5.79 (0.42)
44. How realistic was the difficulty of navigating the colonoscope around mucosal folds?	5.63 (0.50)
9. Looping	
48. How realistic was the ease with which loops formed?	5.74 (0.45)
49. How realistic was the extent of any looping that occurred?	5.79 (0.42)
50. During looping, how realistic was the extent of any paradoxical scope motion?	5.63 (0.50)
51. During looping, how realistic was the feel of resistance to movement of the colonoscope shaft?	5.95 (0.23)
52. During looping, how realistic was the location within the colon of resistance and paradoxical motion?	5.79 (0.42)
53. How realistic was the response of the simulator to loop reduction with typical techniques?	5.84 (0.37)
54. Overall, how realistic was the simulation of looping during insertion?	5.74 (0.45)
10. Patient Discomfort	
55. How realistic was the amount of pain/discomfort experienced by the simulated patient?	4.89 (1.05)
56. How realistic was the auditory feedback provided about patient pain/discomfort?	4.21 (1.03)
57. How realistic was the visual feedback provided about patient pain/discomfort?	3.89 (0.99)
58. Overall, how realistic was the simulation of patient pain/discomfort?	4.53 (1.17)
N/A	
59. Overall, how realistic was the colonoscopy simulator?	N/A