

The efficacy of training insertion skill on a physical model colonoscopy simulator

Authors

Annaliese M. Plooy¹, Andrew Hill^{2,3}, Mark S. Horswill², Alanna St.G. Cresp^{1,4}, Rozemary Karamatic³, Stephan Riek¹, Guy M. Wallis¹, Robin Burgess-Limerick¹, David G. Hewett^{3,4}, Marcus O. Watson^{2,3,4}

Institutions

¹ School of Human Movement Studies, The University of Queensland, St Lucia, Brisbane, Australia

² School of Psychology, The University of Queensland, St Lucia, Brisbane, Australia

³ Clinical Skills Development Service, Metro North Hospital and Health Service, Herston, Brisbane, Australia

⁴ School of Medicine, The University of Queensland, Herston, Brisbane, Australia

submitted 21. March 2016
accepted after revision
29. July 2016

Bibliography

DOI <http://dx.doi.org/10.1055/s-0042-114773>
Published online: 30.9.2016
Endoscopy International Open 2016; 04: E1252–E1260
© Georg Thieme Verlag KG
Stuttgart · New York
E-ISSN 2196-9736

Corresponding author

Andrew Hill, PhD
Clinical Skills Development
Service, Metro North Hospital
and Health Service
School of Psychology
The University of Queensland
St Lucia QLD 4072
Australia
Phone: +61-7-3646 6500
Fax: +61-7-3646 6501
Andrew.Hill@health.qld.gov.au
a.hill@psy.uq.edu.au

Background and study aims: Prior research supports the validity of performance measures derived from the use of a *physical model* colonoscopy simulator – the *Kyoto Kagaku Colonoscope Training Model* (Kyoto Kagaku Co. Ltd., Kyoto, Japan) – for assessing insertion skill. However, its use as a training tool has received little research attention. We assessed the efficacy of a brief structured program to develop basic colonoscope insertion skill through unsupervised practice on the model.

Participants and methods: This was a training study with pretesting and post-testing. Thirty-two colonoscopy novices completed an 11-hour training regime in which they practiced cases on the model in a colonoscopy simulation research laboratory. They also attempted a series of test cases before and after training. For each outcome measure (completion rates, time to cecum and peak force applied to the model), we compared

trainees' post-test performance with the *un-trained novices* and *experienced colonoscopists* from a previously-reported validation study.

Results: Compared with untrained novices, trained novices had higher completion rates and shorter times to cecum overall ($P < .001$), but were out-performed by the experienced colonoscopists on these metrics ($P < .001$). Nevertheless, their performance was generally closer to that of the experienced group. Overall, trained novices did not differ from either experience-level comparison group in the peak forces they applied ($P > .05$). We also present the results broken down by case.

Conclusions: The program can be used to teach trainees basic insertion skill in a more or less self-directed way. Individuals who have completed the program (or similar training on the model) are better prepared to progress to supervised live cases.

Introduction

Effective simulation-based training in colonoscopy insertion skill has the potential to improve patient safety and comfort by reducing the inherent risks associated with procedures performed by trainees under the traditional Halstedian apprenticeship model [1–4]. In this paper, we investigate the extent to which individuals with no prior colonoscopy experience can learn to insert the colonoscope to cecum both efficiently and safely via a relatively brief training regime using a physical model simulator.

There is published evidence that insertion skill acquired through practice on *virtual reality* colonoscopy simulators can transfer to the clinical environment for two such devices: the *Endoscopy Accutouch System* [5–7] (Immersion Medical; Gaithersburg, MD) and the *GI Mentor II* [8,9] (Simbionix Corp. USA; Cleveland, OH). However, the associated performance advantages tend to be short-lived [2,5,8,10,11], perhaps in part

because these devices offer relatively unrealistic approximations of key aspects of the task relevant to insertion, such as looping [12].

In contrast, two *physical model* simulators, 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), have been shown to simulate looping more realistically [12]. Further, evidence from several recent studies supports the construct validity of several performance measures derived from use of the Kyoto Kagaku model for the assessment of insertion skill [13–15]. The first of these studies showed that, compared with experienced colonoscopists, novices had lower completion rates, took longer to reach the cecum, and (for 2 of the 4 colon cases tested) exerted more force on the colon model [13], mirroring experience-related differences found in real colonoscopy [16–22]. A subsequent study found comparable experienced-novice differences when a magnetic endoscopic imaging device was used in con-

License terms



junction with custom software to automate measurement of the colonoscope's progression through the synthetic colon [14]. The final study replicated the experienced-novice differences in procedure time, and also demonstrated experience-level effects for a novel suite of observational metrics used in conjunction with the model [15]. Such comparisons between user-groups known to have differing levels of experience are a common means of generating evidence for the construct validity of metrics associated with the use of a simulation device; that is, the replication of real-world performance differences implies that the simulation taps into the skill that the metrics purport to measure [23]. Despite these promising findings, studies that have evaluated the use of the Kyoto Kagaku model as a training tool (rather than an assessment tool) have produced mixed results [9,24]. One recent study found that surgical residents who engaged in *unstructured* training using the model showed no improvement in their Global Assessment of Gastrointestinal Endoscopic Skills scores from pretest to posttest [9]. In contrast, the results of another study suggest that individualized training under the direct supervision of an attending endoscopist using the Kyoto Kagaku model as a substitute for real patients may lead to faster, more effective cecal intubation [24]. Given the paucity and limitations of existing empirical evidence, a need for further research has been identified [24,25]. The present study is the first to assess the efficacy of using a *structured* training program that does not require the presence of an experienced endoscopist to develop colonoscopy insertion skill through practice on the model.

Participants and methods



Colonoscopy novices completed an unsupervised, structured insertion skill training program using a physical model simulator, and were assessed before and after training. To examine the efficacy of the training method, we compared the novices' performance on four cases at post-test with that of untrained novices and experienced colonoscopists from a study reported previously that validated performance metrics derived from use of the model (i.e., completion to cecum rate, segment completion rate, time to cecum, and peak forces applied to the model) [13]. In addition, we compared the pre-training and post-training performance of the trainees on these same metrics. As well as potentially capturing training-related decreases in the use of force, a key rationale for including force measurements was to address the potential risk that unsupervised trainees might improve their completion rates and times simply by pushing harder. The research was approved by the Human Research Ethics Committees of the Royal Brisbane and Women's Hospital and The University of Queensland. All participants gave informed consent prior to participating.

Participants Novice trainees

The *novice trainees* who participated in the current study were 32 first-year medical students with no prior colonoscopy experience. (Assuming that, in the population, the average novice would experience a moderate training effect of 0.5 SD improvement on each outcome measure, power analysis indicated that at least 26 trainees were required for an 80% probability of detecting these effects.) Trainees were recruited and tested between June 2009 and April 2010, and paid AU220 dollars compensation for their time and travel expenses.



Fig. 1 The Kyoto Kagaku Colonoscope Training Model, without its abdomen cover.

Experience-level comparison groups

These 2 groups participated in a separate, previously-published validation study involving the Kyoto Kagaku simulator [13]. Here, we use the data from that study to provide benchmarks against which to judge the trainees' performance. On average, the 21 *experienced colonoscopists* had 12.10 years of experience in endoscopic practice (range 2–35; $SD=9.41$), including a mean of 9,798 colonoscopies (range 800–40,500; $SD=11,751$). The 18 *untrained novices* were first- and second-year medical students.

Physical model colonoscopy simulator

The *Kyoto Kagaku Colonoscope Training Model* (Kyoto Kagaku Co. Ltd, Kyoto, Japan) comprises a life-size molded plastic torso with a synthetic colon mounted inside (Fig. 1). The colon is tethered to the torso by a series of rubber rings connected (either directly, or via springs) to Velcro-backed fixtures. The model comes with layout guides for 6 standard case configurations, of which we used four to assess insertion skill. These were a relatively straightforward *introductory case* (Case 2), and 3 cases in which loops could not be avoided: an *alpha loop case* (Case 3); a *reverse alpha loop case* (Case 6); and an “N” loop case that also included a drooping transverse colon (Case 4). In addition, we used three modified cases in the study. The *orientation case* (Case 1A) was a version of Case 1 with the rectum stretched out to make insertion easier. Training sessions also employed: a *modified “N” loop case* (Case 4A), which was essentially Case 4 with a straightened transverse colon; and a *modified alpha loop case* (Case 5A), with a deeper transverse than the standard Case 5. The model was always set up and lubricated as per the manufacturer's instructions.

Additional equipment

Participants used an Olympus endoscopy system (Exera II CLV-180 light source and CV-180 processor; OEV203 monitor and CF-H180DL colonoscope; Olympus Medical Systems Corp., Tokyo, Japan) for all sessions. The colon model was supported by a height-adjustable table, and presented in the supine position with a transparent plexiglass sheet in place of its abdomen cover. During test sessions, a video camera recorded the progress of the colonoscope through the model (the light emitted at the tip was visible through the synthetic mucosa). In addition, we used a removable custom-made plastic barrier (the *abdominal occluder*) that could prevent the participant from seeing inside the model without obstructing the camera's view.

As in the previously-reported validation study [13], a force plate (FP4060-NC, Bertec Corporation, Columbus, OH) was interposed between the table and the model to measure force applied to the model in the directions x, y, and z. The x-axis of the force plate was aligned with the model's superior-inferior axis. The force plate was connected to a laptop computer via an analogue-to-digital data acquisition card (DAQ USB-6229 BNC, National Instruments, Austin, TX), and the laptop ran custom software developed in LabView 7.01 (National Instruments, Austin, TX), which sampled force data at 100 Hz. The force plate was surrounded (with a clearance perimeter of 5 mm) by a foam rubber block that was covered with a hospital bed-sheet. The force plate was concealed under a liquid-absorbent underpad, and thin sheets of non-slip rubber with very low force absorptive qualities prevented the model from sliding.

Training study procedure

Sessions were conducted in a university simulation research laboratory, overseen by a research technician. The arrangement of equipment mimicked a procedure room as per the previously-reported validation study [13]. The study comprised 18 1-hour sessions (2 sessions \times 9 weeks). There were 4 stages: orientation (1 session); pretest (2 sessions); training (11 sessions); and post-test (4 sessions). The colon case used varied from session-to-session (as detailed below), but participants were not told which case they would be completing. Before each session, a research technician adjusted the table to the participant's preferred height.

Orientation (Session 1)

Trainees were introduced to colonoscopy and basic colonoscope operation via instructional videos and one-on-one instruction by a research technician. This included basic colorectal anatomy and how to hold the colonoscope, use the controls, manipulate the tip, and torque steer. Finally, trainees practiced on the *orientation case* for 20 minutes with the abdominal occluder in place. This session was equivalent to the preparations undergone by untrained novices in the validation study prior to testing [13].

Pretest (Sessions 2 & 3)

In each session, trainees made 2 attempts (maximum 20 minutes each, separated by a 5-minute break) to complete a case with the abdominal occluder in place (Session 2, *introductory*; Session 3, *alpha loop*). At the beginning of Session 2, a video explained the general procedures for test sessions and instructed participants to treat the colon model as though it were a real patient.

Training (Sessions 4 to 14)

Each session comprised 2 20-minute practice blocks on 1 case, separated by a 5-minute break (Sessions 4 to 6, *alpha loop*; Ses-

sions 7 and 8, *modified "N" loop*; Sessions 9 and 10, *modified alpha loop*; Sessions 11 and 12, *reverse alpha loop*; and Sessions 13 and 14, *"N" loop*). Participants who completed the case in under 20 minutes were allowed another insertion attempt (timing was suspended while the colonoscope was removed and the case reset by a research technician).

The training regime also incorporated an exploratory experimental manipulation: Participants were randomly assigned to 1 of 2 training protocols (16 participants each): *standard visual feedback* or *augmented visual feedback*. For the standard visual feedback group, the abdominal occluder remained in place throughout training. For the augmented visual feedback group, it was removed by a research technician after the first 5 minutes of each practice block, allowing the trainee direct visual access to the colon.

A video shown at the beginning of Session 4 explained the general training procedures for the appropriate trainee group and encouraged participants to experiment with different techniques during training sessions. Additional videos presented at the beginning of training sessions covered the following topics: sigmoid loop reduction (Sessions 4 and 5); common errors (Sessions 9 and 10); navigating through the descending colon (Sessions 9 and 10); and further loop reduction tips (Session 11).

Post-test (Sessions 15 to 18)

These sessions were conducted as per the pre-test sessions. The cases were: Session 15, *reverse alpha loop*; Session 16, *"N" loop*; Session 17, *alpha loop*; and Session 18, *introductory*. These same cases were attempted under similar conditions by the untrained novices and experienced colonoscopists in the validation study [13].

Data scoring

Completion to cecum and segment completion

A researcher examined the video-recording of each test procedure to determine whether the colonoscope tip had reached the end of each anatomical segment of the model (rectum, sigmoid, descending, transverse, and ascending) within the 20-minute time limit. For measurement reliability, a participant was scored as having "completed to cecum" on a particular case if they reached the end of the colon on both attempts. (Note: The model does not have a cecum as such.) Because the "all or none" nature of the *completion to cecum* measure makes it insensitive to incremental improvements in a trainee's ability to advance the scope through the colon, we also calculated the participant's *segment completion* score for each case, which was the average number of anatomical segments completed, expressed as a percentage.

Time to cecum

This was defined as the time elapsed between the tip of the scope passing the anal verge and reaching the end of the colon, and was calculated by a researcher using the video time-code. If the participant failed to complete the procedure, we took the maximum time allowed (20 minutes) as their completion time. For each test case, we averaged *time to cecum* across each subject's 2 attempts, to maximize measurement reliability.

Peak force

The data were filtered using a second-order, dual-pass Butterworth filter (low pass, $f_0 = 5$ Hz) to remove high-frequency noise via custom LabVIEW 7.1 (National Instruments, Austin, TX) software. For each procedure, the software extracted the maximum

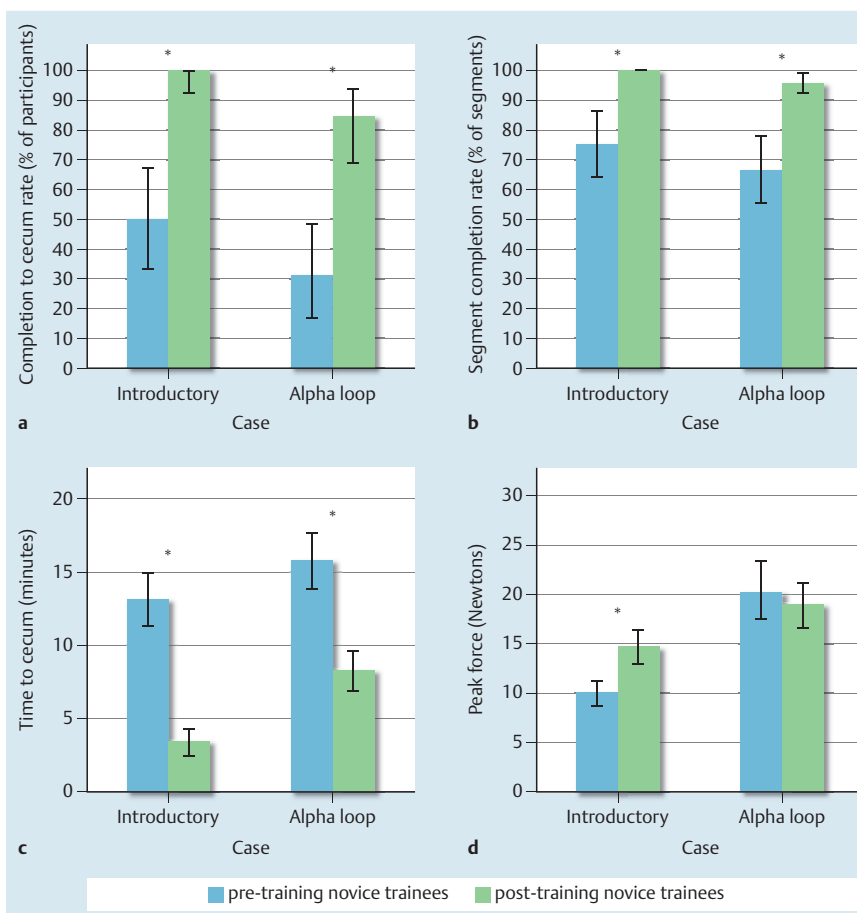


Fig. 2 **a** Overall completion rate, **b** mean segment completion rate, **c** time to cecum and **d** peak force for novice trainees before and after training, arranged by case (with 95% confidence intervals). Asterisks between bars indicate significant pre-post differences (all p 's < .001; see Table 1 for values). Note that Jeffreys confidence intervals are shown for the completion to cecum rates due to the binomial nature of the data [26].

Outcome Measure	Statistical test (and measure of effect size)	P value (and effect size)	
		Introductory case	Alpha loop case
Completion to cecum rate (% of participants)	McNemar's test (w) ¹	< .0001 (2.82)	< .0001 (2.76)
Segment completion rate (% of segments)	t -test (Cohen's d) ²	.0001 (-1.09)	< .0001 (-1.24)
Time to cecum (minutes)	t -test (Cohen's d) ²	< .0001 (2.37)	< .0001 (1.53)
Peak force (Newtons)	t -test (Cohen's d) ²	< .0001 (-1.02)	.3797 (0.20)

¹ For w , values of .50 or greater may be regarded as indicating a large effect [27].

² Cohen's d = the difference between means in units of pooled standard deviation [28] ($\pm .20$ = small; $\pm .50$ = medium; $\pm .80$ = large).

Table 1 Results of statistical tests comparing novice trainees before and after training for each outcome measure by case.

force (in Newtons) applied to the model in the "push" (superior) and "pull" (inferior) directions during insertion, which we averaged across each subject's two attempts at each case to improve measurement reliability.

Statistical analyses

For all analyses, we used IBM SPSS Statistics 19 (SPSS Inc., Chicago, IL). Alpha was set at .05. Preliminary analyses (detailed in the supplement) revealed no significant difference between the *standard visual feedback* and *augmented visual feedback* groups on any of the outcome measures, and power analyses indicated that the groups performed so similarly that, across measures, up to 9894 trainees would be required for an 80% probability of detecting a significant difference. Consequently, all substantive analyses were conducted with the 2 trainee groups combined.

Novice trainees pre- vs. post-training

To directly assess training effects, we compared the pre- and post-test performance of trainees for each of the 2 cases that they attempted at both times. We used McNemar's test to assess changes in the completion to cecum rate, and paired samples t -tests to detect pre-post differences in each of the other outcome measures.

Post-training novice trainees vs. untrained novices and experienced colonoscopists

These analyses were focused on comparing post-training novice trainees with separate groups of (a) untrained novices and (b) experienced colonoscopists, respectively, for each outcome measure. However, we did not compare the untrained novices with experienced colonoscopists (for these analyses, see the previously-published validation study [13]). Rather, we used the validation study participants exclusively as experience-level comparison groups to assess the efficacy of the training program.

Table 2 Means, standard deviations and ranges for each outcome measure by group (averaging over all four cases), and results of statistical tests comparing post-training novice trainees with untrained novices and experienced colonoscopists, respectively.

Outcome Measure	Mean (SD)			Range			Statistical test (and measure of effect size)	P value (and effect size)	
	Untrained novices ¹	Post-training novice trainees	Experienced colonoscopists ¹	Untrained novices ¹	Post-training novice trainees	Experienced colonoscopists ¹		Untrained novices vs. post-training novice trainees	Post-training novice trainees vs. experienced colonoscopists
Completion to cecum rate (% of cases)	21 (N/A)	66 (N/A)	87 (N/A)	N/A	N/A	N/A	t-test (Cohen's d) ²	<.0001 (−2.18)	<.0002 (1.11)
Segment completion rate (% of segments)	54 (49)	88 (30)	97 (16)	13–95	55–100	85–100	F-test (η^2) ³	<.0001 (.62)	<.0001 (.26)
Time to cecum (minutes)	17 (8)	11 (7)	5 (7)	9–20	5–18	2–12	F-test (η^2) ³	<.0001 (.61)	<.0001 (.63)
Peak force (Newtons)	20 (16)	19 (13)	18 (11)	7–37	9–36	10–30	F-test (η^2) ³	.4116 (.01)	.3725 (.02)

¹ These two groups participated in a separate, previously-published validation study [13]. Data from that study are used here to provide benchmarks against which to judge the novice trainees' performance after training.

² Cohen's d = the difference between means in units of pooled standard deviation [28] (± 20 = small; $\pm .50$ = medium; $\pm .80$ = large) [27].

³ η^2 = the proportion of between-groups variance explained (.01 = small; .06 = medium; .14 = large) [29].

Overall completion to cecum rates (averaged over the four post-test cases) were compared in two independent groups *t*-tests (one for each experience-level comparison). Similarly, for segment completion rate, time to cecum and peak force, we used a pair of 2 (group) \times 4 (case) mixed-model ANOVAs to assess group differences in overall performance (i.e., the main effect of group). For each outcome measure, we also quantified the overall percentage of improvement in post-training novice trainees along the continuum from untrained novice to experienced colonoscopist, defining the untrained novice group's performance as the baseline (0%) and the experienced colonoscopists' performance as the end-point (100%).

We followed-up the overall performance analyses by comparing the post-training novice trainees with each experience-level comparison group for each post-test case. For completion to cecum rates, we used a series of Fisher's exact tests to compare the groups on the percentage of participants who completed the relevant case. For the remaining outcome measures, we used independent groups *t*-tests to compare group means, substituting Welch's *t*-test when group variances differed significantly. Note that, although this analysis strategy meant that each set of post-training novice trainee data was included in two tests, we did not adjust for multiple comparisons. To do so would arguably have been less conservative in relation to the comparison with the experienced colonoscopists because it would have increased the probability of incorrectly concluding that the training had made the trainees' performance indistinguishable from that of the experienced group.

Results

Novice trainees pre- vs. post-training

For the cases that trainees completed both before and after training (i.e. the *introductory case* and the *alpha loop case*), the intervention had a significant and substantial effect on all four outcome measures, with a single exception (i.e. the *peak force* that trainees applied during the *alpha loop case*). **Fig. 2** presents means and confidence intervals for each outcome measure for

these two cases before and after training. Each asterisk indicates a significant training effect. **Table 1** contains the corresponding statistical test results and effect sizes [27,28].

Post-training novice trainees vs. untrained novices and experienced colonoscopists

Overall, post-training novice trainees significantly outperformed untrained novices (and were outperformed by experienced colonoscopists) on 3 of the 4 outcome measures: completion to cecum rate; segment completion rate; and time to cecum. **Table 2** presents overall group means, standard deviations, ranges, results of statistical tests and effect sizes for each outcome measure [27–29]. In sum, the trained novices progressed to 67.79% of experienced performance for completion to cecum rate, 78.55% for segment completion, and 52.00% for time to cecum.

The data for each individual post-test case are illustrated in **Fig. 3**, which presents group means and confidence intervals for each outcome measure. Each asterisk indicates a significant difference between the post-training novice trainee group and the adjacent experience-level comparison group. **Table 3** contains the corresponding statistical test results and effect sizes [27,28,30].

Discussion

To our knowledge, this study is the first to demonstrate the efficacy of using a structured training program without supervision from an experienced endoscopist to develop colonoscopy insertion skill through practice on a commercially-available physical model colonoscopy simulator (as opposed to the more expensive and cumbersome virtual reality simulators currently available, which simulate looping less realistically and provide relatively poor simulation of natural haptic feedback [12,31]). After participating in 11 one-hour training sessions, novice trainees had significantly higher overall *completion to cecum* rates and *segment completion* rates than untrained novices, and their overall *time to cecum* was significantly shorter, with most case-level comparisons also indicating large, significant effects. Analyses conducted

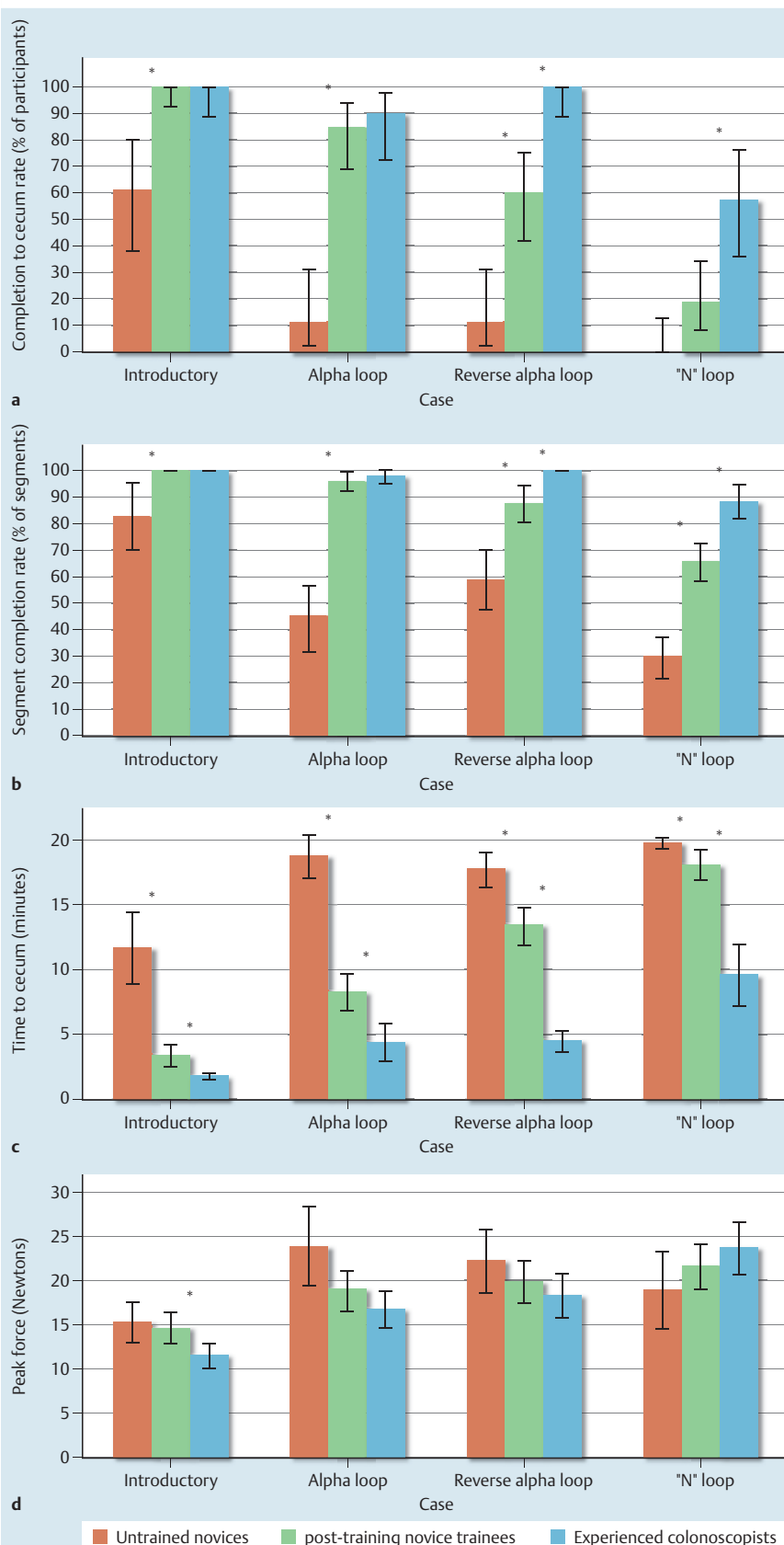


Fig. 3 **a** Overall completion to cecum rate, **b** mean segment completion rate, **c** time to cecum and **d** peak force for each participant group, arranged by case (with 95% confidence intervals). Asterisks between bars indicate significant differences between post-training novice trainees and the adjacent group ($P < .02$; see [Table 3](#) for values). Note that Jeffreys confidence intervals are shown for the completion to cecum rates due to the binomial nature of the data [26].

Table 3 Results of statistical tests (and effect sizes) comparing post-training novice trainees with untrained novices and experienced colonoscopists, respectively, for each outcome measure by case.

Outcome Measure	Comparison	Statistical test (and measure of effect size)	P value (and effect size)			
			Introductory case	Alpha loop case	Reverse alpha loop case	"N" loop case
Completion to cecum rate (% of participants)	Untrained novices vs. post-training novice trainees	Fisher's exact test (<i>phi</i> coefficient) ¹	.0003 (.54)	<.0001 (.72)	.0010 (.47)	.0754 (.28)
	Post-training novice trainees vs. experienced colonoscopists	Fisher's exact test (<i>phi</i> coefficient) ¹	N/A ² (N/A) ²	.6897 (.09)	.0006 (.46)	.0069 (.40)
Segment completion rate (% of segments)	Untrained novices vs. post-training novice trainees	<i>t</i> -test (Cohen's <i>d</i>) ³	.0169 ⁴ (−1.05)	<.0001 ^c (−2.84)	<.0001 (−1.40)	<.0001 (−1.95)
	Post-training novice trainees vs. experienced colonoscopists	<i>t</i> -test (Cohen's <i>d</i>) ³	N/A ² (N/A) ²	.3710 (−0.23)	.0018 ^c (−0.77)	<.0001 (−1.21)
Time to cecum (minutes)	Untrained novices vs. post-training novice trainees	<i>t</i> -test (Cohen's <i>d</i>) ³	<.0001 ⁴ (2.01)	<.0001 (2.68)	.0003 (1.15)	.0071 ⁴ (0.65)
	Post-training novice trainees vs. experienced colonoscopists	<i>t</i> -test (Cohen's <i>d</i>) ³	.0015 ⁴ (0.81)	.0007 (1.02)	<.0001 ⁴ (2.57 ^d)	<.0001 ⁴ (1.99)
Peak force (Newtons)	Untrained novices vs. post-training novice trainees	<i>t</i> -test (Cohen's <i>d</i>) ³	.6659 (0.13)	.0663 ^c (0.63)	.2642 (0.33)	.2782 (−0.32)
	Post-training novice trainees vs. experienced colonoscopists	<i>t</i> -test (Cohen's <i>d</i>) ³	.0158 (0.70)	.1832 (0.38)	.3925 (0.24)	.2794 (−0.31)

¹ The *phi* coefficient = the degree of association between group membership and completion to cecum (.20 to .39 = moderate; .40 to .59 = relatively strong; .60 to .79 = strong; .89 to 1 = very strong) [30].

² All participants in both groups completed to cecum.

³ Cohen's *d* = the difference between means in units of pooled standard deviation [28] (±.20 = small; ±.50 = medium; ±.80 = large) [27].

⁴ Group variances were significantly different, so Welch's *t*-tests are reported.

on data from the cases performed by trainees both before and after training corroborated the large training effects for these outcome measures. Although, unsurprisingly, the trainees were still out-performed by experienced colonoscopists on all three measures at post-test, their performance was generally closer to that of the experienced group than the untrained novices (with the exception of the challenging "N" loop case), as indicated by the relative magnitude of the effect sizes.

It should also be noted that there were considerable individual differences in performance on all outcome measures (as evidenced by the ranges presented in Table 2), with the best-performing trainees equaling or exceeding the performance of some of the experienced colonoscopists after training. This overlap could be seen as indicating that the model is only sufficiently challenging for use in the very early stages of colonoscopic skill acquisition. However, the substantial individual differences among the experienced colonoscopists argue against the ceiling effects that we would expect to see if this were the case for all individuals. An alternative explanation is that basic colonoscope handling and insertion depend primarily on acquiring a specific set of motor skills [32] and that, like all fine motor skills, they are more easily acquired and developed by some individuals than by others [33].

Overall, the trainees did not differ significantly from the experienced colonoscopists in their use of force at post-test, suggesting that improvements in the other outcome measures were not achieved simply by pushing harder (i.e. the trainees did not adopt a blanket strategy of trying to intubate by applying excessive force). However, for the *introductory case* (1 of the 2 cases that yielded a significant experienced-novice difference in the original validation study), our trainees did significantly increase the force that they used at post-test, exceeding that applied by the experienced group. One potential explanation for this is that the lack of looping made the application of additional force a tenable strategy for completing that particular case, but not the

others. Nevertheless, the level of force that the trainees applied after training did not significantly exceed that applied by the untrained novices in the validation study. Hence, this finding should not be over-interpreted and is not in itself evidence that the trainees developed "bad habits". However, it must be acknowledged that, given the unsupervised nature of their training at this early stage in the learning curve, the trainees may have acquired more subtle "bad habits" in their colonoscope handling technique, which a clinical instructor would have corrected. Future research on this training program should therefore also include qualitative performance measures to ascertain whether this is the case. If so, the program may need to be modified to include a small amount of periodic instructor feedback to ensure that poor technique does not become ingrained early on.

Although the primary limitation of our study is that transfer to real patients remains to be demonstrated, the results suggest that individuals who have completed the program (or similar structured training on the model) are better prepared to progress to supervised live cases. That is, it is reasonable to assume that the technical skills acquired by the trainees – which involve learning how to control the colonoscope and how it reacts – are highly likely to transfer to real patient cases (although trainees will still need to acquire skills in other components of colonoscopy competency, [32] such as diagnostic skill, and further develop their colonoscope handling and insertion skills). Hence, this work was a valuable intermediate step and indicates that future studies investigating transfer of training from *structured*, unsupervised training on the Kyoto Kagaku model to real colonoscopy would be worthwhile (noting that existing research evidence calls into question the efficacy of *unstructured* training on the model [9]). However, it must also be acknowledged that, because all of our trainees participated in 2 pretest sessions prior to training, their use of the simulator during these sessions may have contributed to the training effects that were observed at post-test. Therefore, in future implementations of the program that do not involve

pretesting and post-testing, it may be necessary to replace the pre-test with two equivalent sessions of additional practice in order to obtain training effects of the same magnitude. In addition, training dosage effects may be a fruitful avenue for future research in order to determine, for instance, the optimum quantity of simulation-based training to precede a trainee's first real procedure.

Conclusions

The current study has demonstrated that the Kyoto Kagaku Model can be used in conjunction with a structured program to effectively teach trainees basic insertion skill in a more or less self-directed way before they attempt their first real colonoscopy. Compared with other simulation-based alternatives, such training also comes at a relatively low cost in terms of expert supervision and/or dedicated equipment (at least in jurisdictions where the same colonoscopes and endoscopy systems can be used with training models and live patients). Hence, as well as potentially reducing the risks to patient safety and comfort associated with real procedures performed by novices [1–4], such training may also decrease the time that experienced endoscopists must devote to teaching rudimentary insertion skills to trainees, whether in the procedure room or via simulation.

Competing interests: None

Acknowledgements

We thank Tabinda Basit, Welber Marinovic, Hannah Morgan, Victor Selvarajah, and Simranjit Sidhu for their valuable research assistance. This research was supported by the Australian Government Department of Health and Ageing. David Hewett was supported by a Sylvia & Charles Viertel Charitable Foundation Clinical Investigatorship. Guy 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.

References

- Sedlack RE. Endoscopic simulation: where we have been and where we are going. *Gastrointest Endosc* 2005; 61: 216–218
- Sedlack RE. Simulators in training: defining the optimal role for various simulation models in the training environment. *Gastrointest Endosc Clin N Am* 2006; 16: 553–563
- Reznick RK, MacRae H. Teaching surgical skills – changes in the wind. *N Engl J Med* 2006; 355: 2664–2669
- Williams CB, Thomas-Gibson S. Rational colonoscopy, realistic simulation, and accelerated teaching. *Gastrointest Endosc Clin N Am* 2006; 16: 457–470
- 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: 33–37
- 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: 205–211
- Ahlberg G, Hultcrantz R, Jaramillo E et al. Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? *Endoscopy* 2005; 37: 1198–1204
- 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: 361–368
- Gomez PP, Willis RE, Van Sickle K. Evaluation of two flexible colonoscopy simulators and transfer of skills into clinical practice. *J Surg Educ* 2015; 72: 220–227
- Sturm LP, Windsor JA, Cosman PH et al. A systematic review of skills transfer after surgical simulation training. *Ann Surg* 2008; 248: 166–179
- Tsuda S, Scott D, Doyle J et al. Surgical skills training and simulation. *Curr Probl Surg* 2009; 46: 271–370
- Hill A, Horswill MS, Plooy AM et al. A systematic evaluation of the realism of four colonoscopy simulators. *Gastrointest Endosc* 2012; 75: 631–640
- Plooy AM, Hill A, Horswill MS et al. Construct validation of a physical model colonoscopy simulator. *Gastrointest Endosc* 2012; 76: 144–150
- Nerup N, Preisler L, Svendsen MBS et al. Assessment of colonoscopy by use of magnetic endoscopic imaging: design and validation of an automated tool. *Gastrointest Endosc* 2015; 81: 548–554
- Preisler L, Svendsen MBS, Nerup N et al. Simulation-based training for colonoscopy: establishing criteria for competency. *Medicine (Baltimore)* 2015; 94: e440
- Cass OW. Objective evaluation of competence: technical skills in gastrointestinal endoscopy. *Endoscopy* 1995; 27: 86–89
- Marshall JB. Technical proficiency of trainees performing colonoscopy: a learning curve. *Gastrointest Endosc* 1995; 42: 287–291
- Spier BJ, Benson M, Pfau PR et al. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc* 2009; 71: 319–324
- Lee SH, Chung IK, Kim SJ et al. An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. *Gastrointest Endosc* 2008; 67: 683–689
- Chak A, Cooper GS, Blades EW et al. Prospective assessment of colonoscopic intubation skills in trainees. *Gastrointest Endosc* 1996; 44: 217–230
- 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
- Appleyard MN, Mosse CA, Mills TN et al. The measurement of forces exerted during colonoscopy. *Gastrointest Endosc* 2000; 52: 237–240
- Fairhurst K, Strickland A, Maddern GJ. Simulation speak. *J Surg Ed* 2011; 68: 382–386
- Kaltenbach T, Leung C, Wu K et al. Use of the colonoscope training model with the colonoscope 3D imaging probe improved trainee colonoscopy performance: a pilot study. *Dig Dis Sci* 2011; 56: 1496–1502
- Yoshida N, Fernandopulle N, Inada Y et al. Training methods and models for colonoscopic insertion, endoscopic mucosal resection, and endoscopic submucosal dissection. *Dig Dis Sci* 2014; 59: 2081–2090
- Brown LD, Cai TT, Das Gupta A. Interval estimation for a binomial proportion. *Stat Sci* 2001; 16: 101–133
- Cohen J. A power primer. *Psychol Bull* 1992; 112: 155–159
- Rosnow RL, Rosenthal R. Computing contrasts, effect sizes, and counter-nulls on other people's published data: General procedures for research consumers. *Psychol Meth* 1996; 1: 331–340
- Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, New Jersey: L. Erlbaum Associates; 1988
- Kotlik JW, Williams HA. The incorporation of effect size in information technology, learning, and performance research. *Inform Technol Learn Perform J* 2003; 21: 1–7
- Sedlack RE. The state of simulation in endoscopy education: continuing to advance toward our goals. *Gastroenterology* 2013; 144: 9–12
- Zupanc CM, Burgess-Limerick R, Hill A et al. A competency framework for colonoscopy training derived from cognitive task analysis techniques and expert review. *BMC Med Ed* 2015; 15: 216
- Edwards WH. Motor learning and control: from theory to practice. Belmont, CA: Wadsworth Cengage Learning; 2011

Supplementary material



Preliminary analyses comparing the two training protocols

Before conducting the substantive analyses, we investigated whether the exploratory training protocol manipulation (*standard visual feedback* vs. *augmented visual feedback*) led to post-training performance differences between the two groups of novice trainees. To this end, we conducted Fisher's exact test to compare the groups on completion to cecum rate for each case, and a 2 (training group) \times 4 (case) mixed-model analysis of variance for each of the other three outcome measures. Across the four cases completed at post-test, there was no significant difference between the two training protocols for the completion to cecum rate (Fisher's exact tests, all P 's $> .05$), segment completion rate ($F(1,30) = .06$, $P = .816$), time to cecum ($F(1,30) = .02$, $P = .902$), or peak force applied ($F(1,30) = 2.50$, $P = .125$). Further, it made no difference to the pattern of results when pre-training performance was controlled for.

To assess the power of these analyses, we calculated the sample size required to have an 80% probability of detecting each effect with alpha set at .05, assuming that the pattern of results found in the study reflected underlying population differences. For the completion to cecum rates, samples of 888, 228, and 360 participants would be required to obtain significance across the alpha loop, reverse alpha loop, and "N" loop cases, respectively. For the introductory case, the groups had identical completion to cecum rates and hence Fisher's exact test would never reach significance, irrespective of the sample size. For the other outcome measures, the sample sizes that would be required for the ANOVA omnibus tests to reach significance were: segment completion rate, 3056; time to cecum, 9894; and peak force, 76.