

# Validation of the VBLaST pattern cutting task: a learning curve study

Ali M. Linsk<sup>1</sup> · Kimberley R. Monden<sup>3</sup> · Ganesh Sankaranarayanan<sup>3</sup> · Woojin Ahn<sup>4</sup> · Daniel B. Jones<sup>1</sup> · Suvranu De<sup>4</sup> · Steven D. Schweitzberg<sup>5</sup> · Caroline G. L. Cao<sup>2</sup>

Received: 9 September 2016 / Accepted: 16 September 2017 / Published online: 19 October 2017  
© Springer Science+Business Media, LLC 2017

## Abstract

**Background** Mastery of laparoscopic skills is essential in surgical practice and requires considerable time and effort to achieve. The Virtual Basic Laparoscopic Skill Trainer (VBLaST-PC<sup>®</sup>) is a virtual simulator that was developed as a computerized version of the pattern cutting (PC) task in the Fundamentals of Laparoscopic Surgery (FLS) system. To establish convergent validity for the VBLaST-PC<sup>®</sup>, we assessed trainees' learning curves using the cumulative summation (CUSUM) method and compared them with those on the FLS.

**Methods** Twenty-four medical students were randomly assigned to an FLS training group, a VBLaST training group, or a control group. Fifteen training sessions, 30 min in duration per session per day, were conducted over 3 weeks. All subjects completed pretest, posttest, and retention test (2 weeks after posttest) on both the FLS and VBLaST<sup>®</sup> simulators. Performance data, including time, error, FLS score, learning rate, learning plateau, and CUSUM score, were analyzed.

**Results** The learning curve for all trained subjects demonstrated increasing performance and a performance plateau. CUSUM analyses showed that five of the seven subjects reached the intermediate proficiency level but none reached the expert proficiency level after 150 practice trials. Performance was significantly improved after simulation training, but only in the assigned simulator. No significant decay of skills after 2 weeks of disuse was observed. Control subjects did not show any learning on the FLS simulator, but improved continually in the VBLaST simulator.

**Conclusions** Although VBLaST<sup>®</sup>- and FLS-trained subjects demonstrated similar learning rates and plateaus, the majority of subjects required more than 150 trials to achieve proficiency. Trained subjects demonstrated improved performance in only the assigned simulator, indicating specificity of training. The virtual simulator may provide better opportunities for learning, especially with limited training exposure.

**Keywords** Learning curve · Cumulative summation (CUSUM) · Virtual reality · Surgical training · Convergent validity

Presented at the SAGES 2017 Annual Meeting, March 22–25, 2017, Houston, Texas.

✉ Caroline G. L. Cao  
caroline.cao@wright.edu

- <sup>1</sup> Beth Israel Deaconess Medical Center, Boston, MA, USA
- <sup>2</sup> Wright State University, 207 Russ Engineering Center, 3640 Colonel Glenn Hwy, Dayton, OH 45435, USA
- <sup>3</sup> Baylor University Medical Center at Dallas, Dallas, TX, USA
- <sup>4</sup> Rensselaer Polytechnic Institute, Troy, NY, USA
- <sup>5</sup> University at Buffalo School of Medicine and Biomedical Sciences, Buffalo, NY, USA

The laparoscopic approach has become the standard of care for a wide variety of surgical procedures and has the advantages of faster recovery, minimal blood loss, and lower cost of treatment [1]. Despite the many benefits of minimally invasive surgery (MIS), the technique is more demanding for surgeons and requires extensive training. This is due to the increased sensorimotor challenges associated with this technique such as hand–eye coordination, two-dimensional field of view, and lack of perceivable haptic feedback [2]. As a result, laparoscopic surgery trainees must undergo a

substantial amount of preparation using simulators prior to performing live operations.

The current standard for basic laparoscopic skill development is the Fundamentals of Laparoscopic Surgery (FLS) curriculum [3, 4], which is administered by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) and the American College of Surgeons (ACS). The FLS trainer is a physical box trainer based on the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) [5]. The five FLS tasks used in the manual skills portion of the curriculum are peg transfer, pattern cutting (PC), ligating loop, suture with intracorporeal knot, and suture with extracorporeal knot. Achieving proficiency in these tasks provides the foundation of laparoscopic surgical skill performance. Since 2009 in the USA, successful completion of the FLS exam is a requirement before being eligible to take the Qualifying Examination of the American Board of Surgery.

Despite being the standard in laparoscopic training, there are major drawbacks to the FLS practical exam, including difficulty in evaluating performance objectively and the time needed to score manually. To overcome these issues, virtual reality (VR)-based simulators can be used to replace physical models. VR simulators enable objective and automated assessment of performance, in real time, and without the need for proctors. Moreover, they permit unlimited training without the expense of consumables. They can also provide haptic feedback, which has already been shown to be an essential component of minimally invasive surgery simulations [6, 7]. VR-based simulators have been shown to transfer effective technical skills to the operating room environment [8, 9]. The Virtual Basic Laparoscopic Skill Trainer (VBLaST) was developed as the VR version of the FLS trainer [10] and maps the five FLS skills to a virtual environment. The VBLaST has shown face validity, as well as construct, concurrent, and discriminant validity [10–13].

The cumulative summation (CUSUM) is a criterion-based method that is commonly used for characterizing learning curves. It is a statistical and graphical tool that analyzes trends for sequential events in time and hence can be used for quality control of individual performance and group performance. It can be applied in the learning phases, such as while learning a new procedure, and at the end of the training phase after the acquisition of the skill [14, 15]. Previous research has examined the learning curve for the VBLaST peg transfer task and found it to be comparable to that of the FLS peg transfer task. The objective of this study was to continue the validation process of the VBLaST simulator and demonstrate the convergent validity of the PC task using the CUSUM method. To demonstrate convergent validity, the system must be at least as effective as a commonly accepted training system, such as the FLS. Therefore, it was expected

that the learning curves on the VBLaST-PC and FLS are similar, with performance improving with practice. In addition, subjects with training on either simulator would perform better than those with no training in the posttest and retention test.

## Materials and methods

### Subjects

Based on prior learning curve studies and power calculations, five subjects were necessary for each of the three conditions in this learning curve study. Thirty medical students were recruited to allow for attrition, which was anticipated due to the long-time commitment required.

An IRB-approved recruitment email was sent to all Tufts University Medical students. Inclusion criteria were as follows: little or no prior experience with surgery or surgical simulators, normal or corrected to normal vision, and no motor impairment that prevented the handling of two laparoscopic tools in the surgical simulators. Subjects were compensated for their participation.

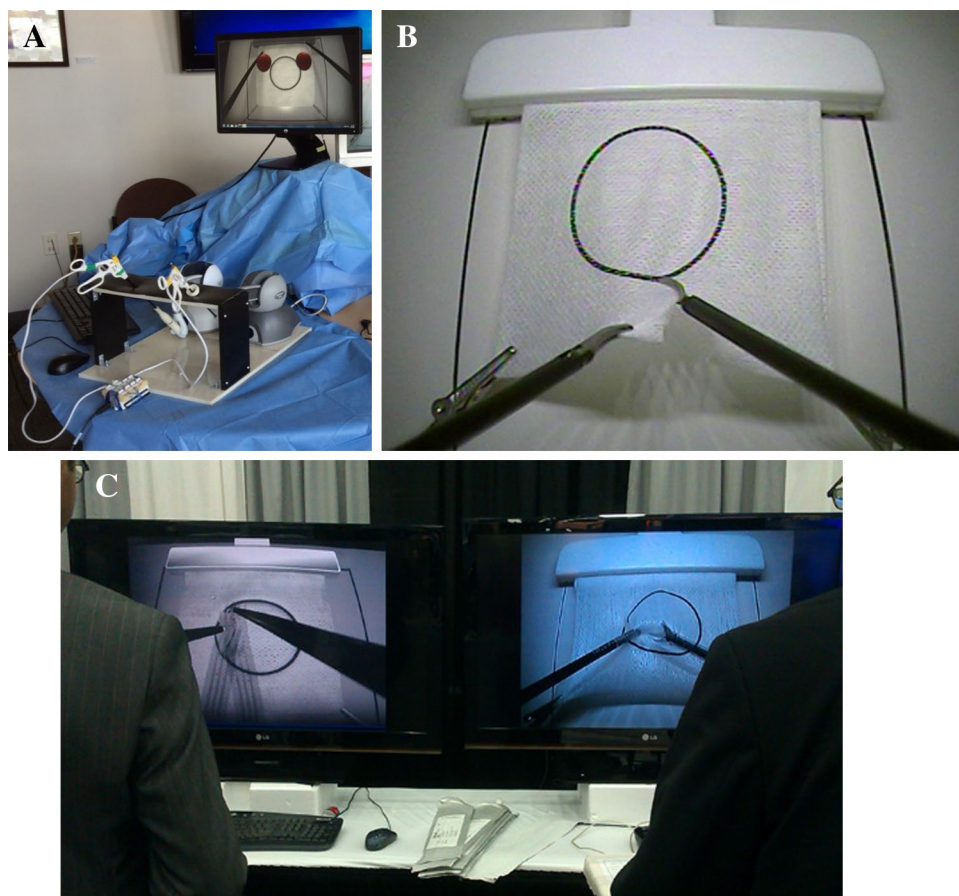
Ten subjects were randomly assigned to each of the three conditions (control, FLS, and VBLaST). At the end of the study, there were nine subjects in the control group, eight in the FLS training group, and seven in the VBLaST training group, due to attrition.

### Equipment

The FLS system (Fig. 1A) used was the standard SAGES-approved trainer box. The view of the task space was filmed with a fixed focal length camera and displayed on a monitor for the subject. A digital capture device (AVerMedia, Milpitas, CA, USA) was used to record subjects' performance inside the task space. The video was used to gather timing and error measurements for data analysis.

The VBLaST-PC system (Fig. 1B) consisted of two laparoscopic tools connected to haptic devices mounted in front of a monitor, and a VR environment simulating the FLS-PC task. Custom-developed computational software in the VBLaST-PC simulated the interaction between the tools and objects in the virtual environment. The two Geomagic Touch haptic devices (3D Systems Inc.) connected to the instrumented tools provided force feedback to the user. Figure 1C shows side-by-side comparison of the two systems. Even though the system is capable of tracking and calculating performance variables such as instrument path length and smoothness, only time to task completion and errors were used in this study to provide a fair comparison to the FLS.

**Fig. 1** **A** VBLaST-PC, **B** FLS-PC (<http://www.flsprogram.org>), and **C** comparison of VBLaST-PC on the left and FLS-PC on the right



### Experimental design and procedure

The PC task was used in this convergent validity study. The FLS-PC task requires the subject to cut a circle out of a 4 cm × 4 cm piece of gauze along a pre-marked black line as quickly and accurately as possible. The task is completed using laparoscopic instruments and the official FLS box trainer as shown below. The same task was completed with the VBLaST<sup>®</sup> system using laparoscopic instruments and computational software.

In this mixed experimental design, participants were randomly assigned to one of three conditions: control, VBLaST, and FLS. Subjects in the control group did not receive any training on the task, while those in the two training groups received training on the assigned simulator over a period of 3 weeks. Demographic data including age, medical school year, and information regarding previous experience and laparoscopic surgery were gathered for all subjects. Before the testing session began, all subjects watched an instructional video that demonstrated the proper procedure to perform the PC task in both the FLS and VBLaST systems. All subjects then performed the task

once using both simulators to establish their own baseline performance. This also served as the pretest assessment. The order of simulators was counterbalanced—half the subjects used the FLS system first for the pretest and half the subjects used the VBLaST system first.

Subjects randomized in the training groups were asked to attend one training session per day, 5 days per week, for three consecutive weeks equaling a total of 15, 30-min sessions. During each 30-min session, subjects were asked to perform ten trials of the PC task or attempt as many trials as possible, whichever was shorter. The experimenter was always present during these training sessions to answer questions and provide instruction when needed. VBLaST-PC time and error data were automatically recorded by the simulator. FLS training scores were computed manually and time was measured using a stopwatch.

At the end of 3 weeks, all subjects (training and control) performed the PC task on both the FLS and VBLaST to record posttest data. To assess retention, a final session was held 2 weeks after the last training session. Table 1 summarizes our experimental design.

**Table 1** Learning curve study timeline

	Week 1	Week 2	Week 3	Week 4	Week 5
Control group	Pretest (day 1), no training	No training	No training posttest (day 15)	No training	No training, retention test (day 16)
Trained groups	Pretest (day 1), five training sessions	Five training sessions (30 min each)	Five training sessions (30 min each), posttest (day 15)	No training	No training, retention test (day 16)

## Data analysis

### Performance data analysis

Using SPSS, the pretest, posttest, and retention test data (time, error, score) were analyzed using a two-way ANOVA mixed design and Tukey's HSD post hoc test, as well as multiple pairwise comparisons with Bonferroni correction. The criterion for statistical significance was set at  $\alpha = 0.05$ . Outliers were removed from the data and replaced with the group means before analysis. On the FLS simulator, one data point on the time measure and four data points on the error measure were considered outliers. On the VBLaST simulator, three data points in time and four data points in error were considered outliers.

### CUSUM analysis

CUSUM analysis was performed by generating CUSUM charts for all the subjects in both the FLS and the VBLaST training groups. For both the training groups, a criterion on the normalized score was established based on the current accepted proficiency score for the FLS-PC task (score = 72) and an intermediate proficiency score (score = 56) calculated based on the average value of the score for all subjects in the first 40 trials. When the computed PC score for each trial for both groups equaled or exceeded the criterion score, it was defined as a 'success' (1), while a lower score was deemed as 'failure' (0). The acceptable failure rate ( $p_0$ ) was set at 5%, and the unacceptable failure rate ( $p_1$ ) was set at 10% ( $2 \times p_0$ ). Type I and type II errors ( $\alpha$  and  $\beta$ ) were set at 0.05 and 0.20, respectively. Based on those parameters, two decision limits ( $h_0$  and  $h_1$ ) and  $s$ , the target value for CUSUM, were calculated for each successive trial. For each 'success,'  $s$  was subtracted from the previous CUSUM score. For each 'failure,'  $1 - s$  was added to the previous CUSUM score. A negative slope of the CUSUM line indicates success, whereas a positive slope suggests failure. This procedure was repeated for each subject on both training groups. Table 2 shows the CUSUM variables for data analysis.

**Table 2** CUSUM criteria score and parameters

Variable	Values
FLS-PC intermediate proficiency score	56
FLS-PC proficiency score	72
VBLaST-PC intermediate proficiency score	56
VBLaST-PC proficiency score	72
$p_0$	0.05
$p_1 = 2 \times p_0$	0.10
$\alpha$	0.05
$\beta$	0.20
$P = \ln(p_1/p_0)$	0.69
$Q = \ln[(1 - p_0)/(1 - p_1)]$	0.05
$s = Q/(P + Q)$	0.07
$1 - s$	0.93
$a = \ln[(1 - \beta)/\alpha]$	2.77
$b = \ln[(1 - \alpha)/\beta]$	1.56
$h_0 = -b/(P + Q)$	-2.09
$h_1 = a/(P + Q)$	3.71

### Learning plateau and learning rate analysis

We calculated the learning plateau and learning rate using the learning curve data of all subjects in both FLS and VBLaST training conditions based on the method of inverse curve fitting. The learning plateau was defined as the asymptote of the fitted curve and learning rate was defined as the number of trials required to reach 90% of the plateau [16].

## Results

Subjects' baseline performance (pretest) scores are summarized as group means in Table 3. Analysis results indicate that the groups were similar at the beginning of the study, even though the differences in their performance scores on the FLS were significant ( $p = .048$ ). No other measures showed a significant difference between groups.

**Table 3** Subjects' baseline performance as measured by the pretest on FLS and VBLaST

Simulator	Training group	Mean time (SD)		Mean error (SD)		Mean score (SD)	
FLS	Control	282.33 (76.84)	n.s.	11.62 (7.06)	n.s.	11.62 (7.06)	$p = .048$
	FLS	233.53 (85.30)		15.34 (6.04)		22.41 (16.70)	
	VBLaST	210.79 (61.70)		10.61 (5.60)		27.49 (21.66)	
VBLaST	Control	312.78 (138.80)	n.s.	47544.71 (18106.46)	n.s.	13.43 (10.08)	n.s.
	FLS	272.60 (115.70)		61120.23 (27923.55)		11.24 (10.40)	
	VBLaST	314.55 (77.56)		57732.77 (22062.74)		6.17 (9.51)	

n.s. not significant

## FLS simulator

### Time to task completion

Analysis of variance results (see Table 4) showed a significant main effect in training condition [ $F(2,21) = 7.749$ ,  $p = .003$ ,  $\eta^2 = .425$ ], and learning effect as measured in pretest, posttest, and retention test [ $F(2,42) = 21.924$ ,  $p < .001$ ,  $\eta^2 = .511$ ]. No significant interaction between training condition and learning effect was found. Tukey's HSD post hoc test showed that FLS-trained group was significantly different from the control group. There was no difference between the control group and the VBLaST-trained group. Post hoc pairwise comparisons, with Bonferroni correction, showed a significant difference between pretest and posttest, suggesting significant improvement over time, and no significant differences between posttest and retention test (Fig. 2A).

### Error

There was a significant learning effect in the error measure [ $F(2,42) = 22.957$ ,  $p < .001$ ,  $\eta^2 = .522$ ], but no difference as a function of training condition. No interaction between the two factors was observed. Post hoc pairwise comparisons, with Bonferroni correction, showed that errors improved significantly from pretest to posttest, but no significant changes were observed from posttest to retention test, suggesting that learning occurred with training, and the learning effect did not decay with time after the training period (Fig. 2B).

### Normalized score

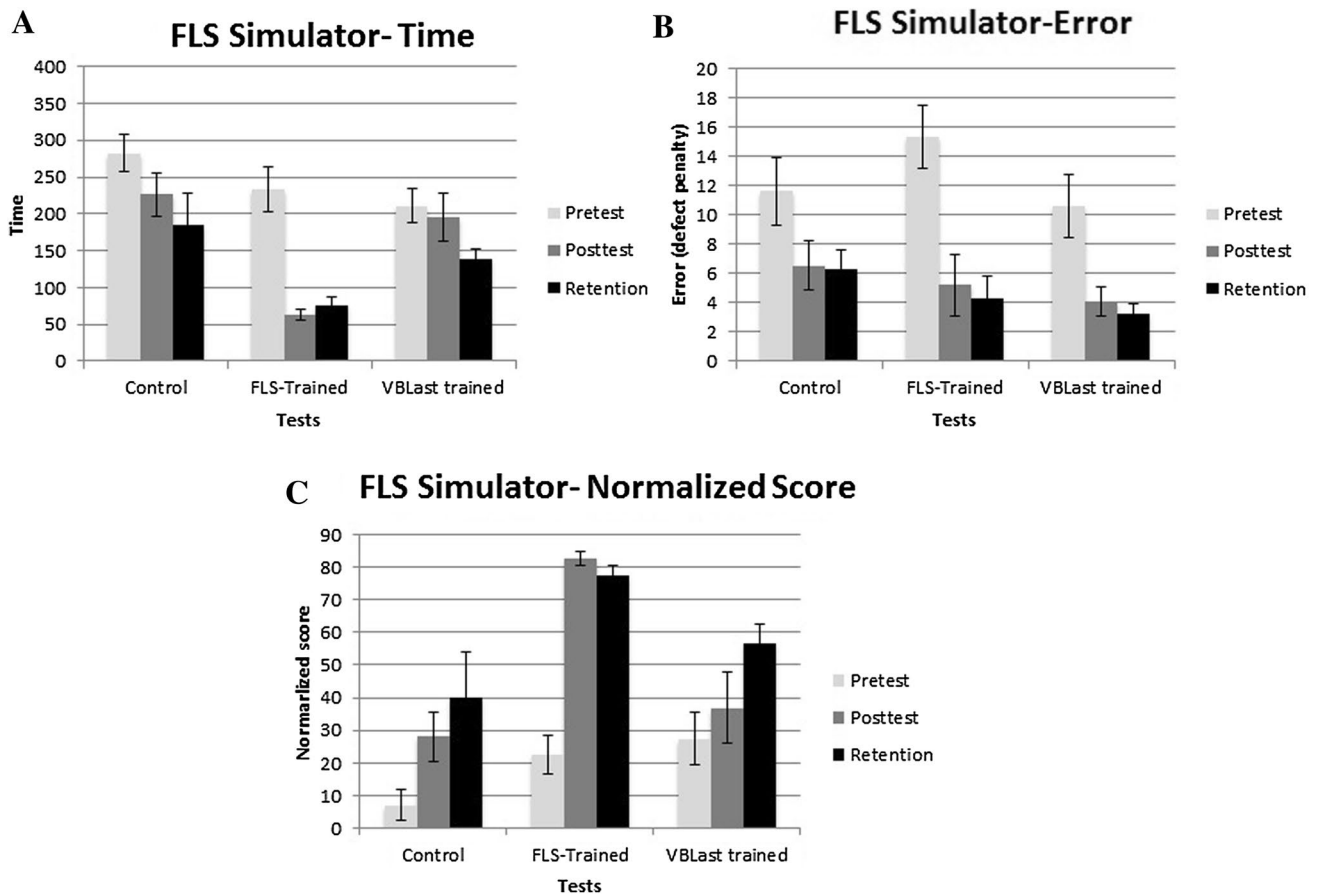
The normalized score showed a significant main effect in the training condition [ $F(2,21) = 42.87$ ,  $p < .001$ ,  $\eta^2 = .803$ ] and a significant learning effect [ $F(2,42) = 40.03$ ,  $p < .001$ ,  $\eta^2 = .656$ ]. There was also a significant interaction between training condition and testing condition [ $F(4,42) = 23.67$ ,  $p < .001$ ,  $\eta^2 = .693$ ]. Tukey's HSD post hoc test showed a significant difference between all training groups. Post hoc pairwise comparisons, with Bonferroni correction, showed that pretest and posttest were significantly different, suggesting learning, but no difference between posttest and retention test scores, indicating no skill decay with time after the training period (Fig. 2C).

### Pretest–posttest

The change in performance from pretest to posttest as an indicator of learning was analyzed. Results from individual  $t$  tests showed that the change in performance for the VBLaST-trained group was not different from that of the control group, whereas the change in performance for the FLS-trained group was significantly different from the control group. This suggests that there was no transfer of learning from the VBLaST environment.

**Table 4** ANOVA results for the 3 × 3 mixed design (3 training conditions × 3 tests)

Simulator	Performance measure	Training condition	$\eta^2$	Tests	$\eta^2$	Training × test interaction	$\eta^2$
FLS	Time (s)	$F(2,21) = 7.749$ , $p = .003$	.425	$F(2,42) = 21.924$ , $p < .001$	.511	$F(4,42) = 3.61$ , $p = .013$	.256
	Error	$F(2,21) = 1.074$ , $p = .360$	.093	$F(2,42) = 22.957$ , $p < .001$	.522	$F(4,42) = 1.078$ , $p = .379$	.093
	Score	$F(2,21) = 42.87$ , $p < .001$	.803	$F(2,42) = 40.03$ , $p < .001$	.656	$F(4,42) = 23.67$ , $p < .001$	.693
VBLaST	Time (s)	$F(2,21) = 4.101$ , $p = .031$	.281	$F(2,42) = 42.087$ , $p < .001$	.667	$F(4,42) = 3.373$ , $p = .018$	.243
	Error	$F(2,21) = 1.407$ , $p = .267$	.118	$F(2,42) = 2.773$ , $p = .074$	.117	$F(4,42) = 2.371$ , $p = .068$	.184
	Score	$F(2,21) = 10.362$ , $p = .001$	.497	$F(2,42) = 63.61$ , $p < .001$	.752	$F(4,42) = 8.414$ , $p < .001$	.445



**Fig. 2** Performance group means for the three training groups in pretest, posttest, and retention test on the FLS simulator (the error bars represent standard deviation): **A** completion time, **B** error, and **C** normalized score

### VBLAST simulator

#### Time to task completion

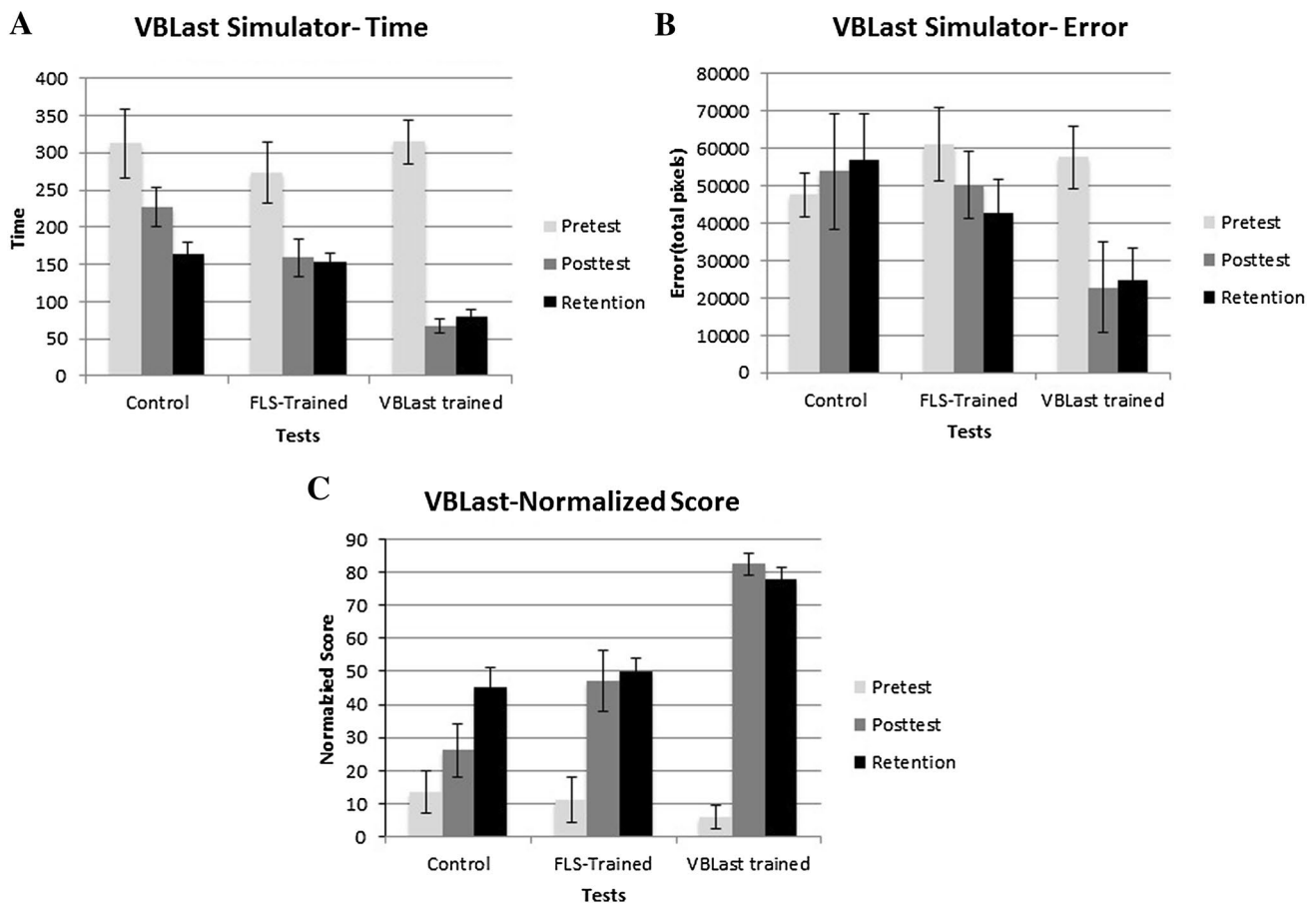
Analysis of variance results (see Table 4) showed a significant difference in subjects' time to task completion as a function of training condition [ $F(2,21) = 4.101$ ,  $p = .031$ ,  $\eta^2 = .281$ ]. There was also a significant learning effect as measured in pretest, posttest, and retention test [ $F(2,42) = 42.087$ ,  $p < .001$ ,  $\eta^2 = .667$ ]. A significant interaction between training and learning was noted [ $F(4,42) = 3.373$ ,  $p = .018$ ,  $\eta^2 = .243$ ]. Tukey's HSD post hoc test showed that the only difference was between the control group and the VBLAST-trained group. There were no significant differences between the other groups, suggesting that FLS skills did not transfer to the VBLAST simulator. Post hoc pairwise comparisons, with Bonferroni correction, showed a significant difference between pretest and posttest, suggesting improvement with training. There were no significant differences between posttest and retention test, suggesting that there was no decay in skill with time after the training period (Fig. 3A).

#### Error

No significant differences were found in any of the factors for the error measure on the VBLAST simulator (Fig. 3B).

#### Normalized score

The normalized score showed a significant main effect in the training factor [ $F(2,21) = 10.362$ ,  $p = .001$ ,  $\eta^2 = .497$ ] and a significant learning effect [ $F(2,42) = 63.61$ ,  $p < .001$ ,  $\eta^2 = .752$ ]. There was a significant interaction between training condition and testing condition [ $F(4,42) = 8.414$ ,  $p < .001$ ,  $\eta^2 = .445$ ]. Tukey's HSD post hoc test showed a significant difference between the subjects in the control and VBLAST-trained groups, and between subjects in the FLS- and VBLAST-trained groups. Control subjects and FLS-trained subjects were not different, suggesting that FLS skills did not transfer to the VBLAST simulator. Again, the pairwise comparisons with Bonferroni correction indicate that significant learning occurred from pretest to posttest, and no significant decay took place between posttest and retention test (Fig. 3C).



**Fig. 3** Performance group means for the three training groups in pretest, posttest, and retention test on the VBLAST-PC simulator (error bars represent standard deviation); **A** completion time, **B** error, and **C** normalized score

#### Pretest–posttest

The change in performance from pretest to posttest as an indicator of learning was analyzed. Results from individual *t* tests showed that the change in performance for the FLS-trained group was not different from that of the control group, whereas the change in performance for the VBLAST-trained group was significantly different from that of the control group. This suggests that there was no transfer of learning from the FLS environment.

#### Cumulative summation analyses for FLS-PC and VBLAST-PC training groups

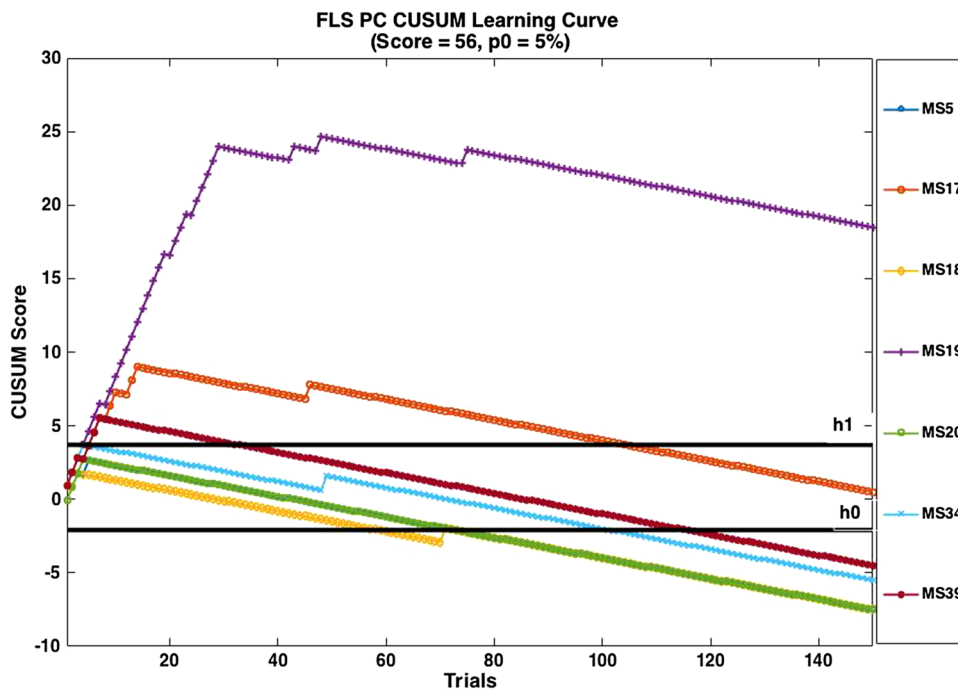
Based on the intermediate proficiency criterion (score = 56) (Fig. 4), five of the seven medical students (MS) achieved the acceptable failure rate of 5% by 150 trials (MS 20, MS 18, and MS 5 at the 73rd trial, MS 34 at the 101st trial, MS 39 at the 115th trial). All subjects showed performance transition points (the trial at which the slope of the CUSUM curve becomes negative),

indicating that they reached the targeted score and were improving as trials progressed (MS 18 at the 3rd trial, MS 20 at the 4th trial, MS 5 at the 5th trial, MS 39 at the 7th trial, MS 17 at the 14th trial, and MS 19 at the 29th trial). The performance of MS 17 was between the two decision limits of  $h_0$  and  $h_1$  and did not reach the acceptable failure rate. MS 39 did not achieve proficiency with acceptable failure rate in the 150 trials.

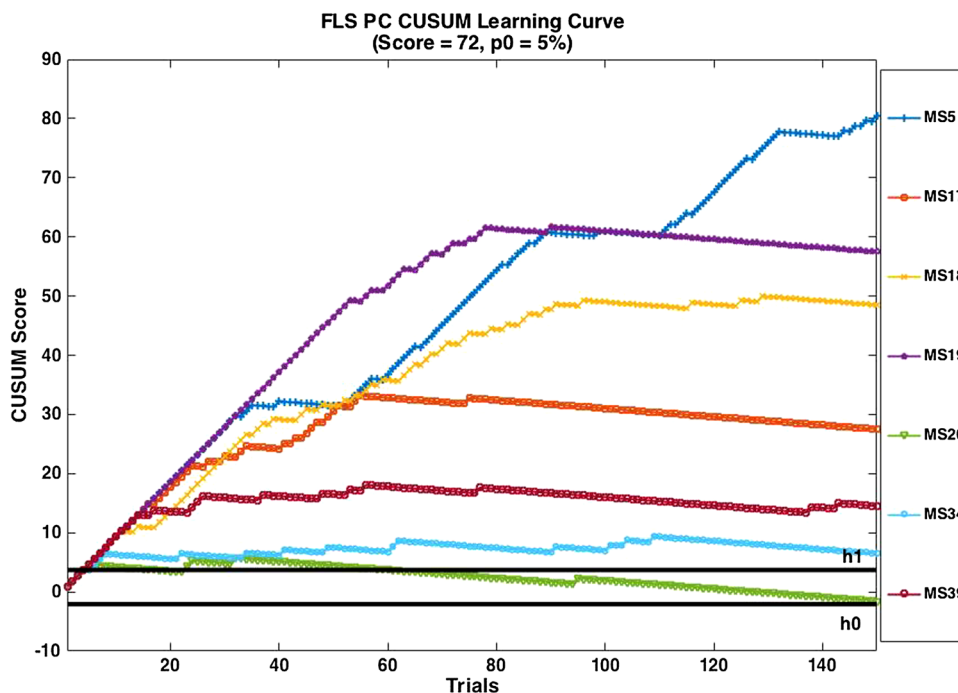
Based on the proficiency score of 72, none of the students achieved the acceptable score rate of 5% (Fig. 5). For one student (MS 20), the performance crossed the upper decision limit  $h_1$  but not the lower decision limit  $h_0$  within 150 trials. Six students showed performance transition points (MS 17 at the 75th trial, MS 39 at the 77th trial, MS 19 at the 90th trial, MS 20 at the 95th trial, and MS 34 at the 109th trial).

Based on the intermediate proficiency criteria score of 52, two students achieved the 5% acceptable failure rate (see Fig. 6) on the VBLAST simulator (MS 37 achieved at the 58th trial and MS 30 at the 143rd trial). The performance of MS 4 was between the two decision limits. All students showed transition points (MS 37 at the 12th trial, MS 4 at

**Fig. 4** CUSUM learning curves for medical students trained on FLS simulator using the intermediate criterion success score of 56, acceptable failure rate  $p_0=5\%$



**Fig. 5** CUSUM learning curves for medical students trained on FLS simulator using the FLS proficiency criterion success score of 72, acceptable failure rate  $p_0=5\%$



the 18th trial, MS 14 at the 32nd trial, Ms 26 and MS 22 at the 49th trial, and MS 14 at the 79th trial).

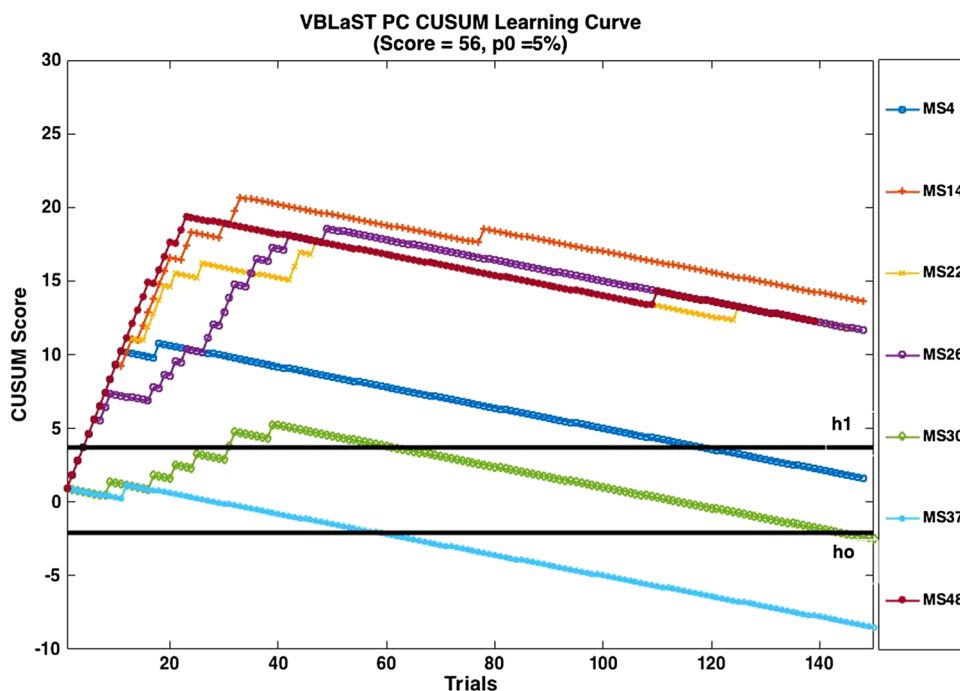
Based on the proficiency criteria score of 72, none of the students achieved the acceptable failure rate of 5% (Fig. 7.) The performance of one student (MS 37) was between the two decision limits  $h_0$  and  $h_1$ . Two medical students showed transition points (MS 30 at the 49th trial and MS 37 at the 41st trial).

**Learning plateau and learning rate for FLS-PC and VBLaST-PC training groups**

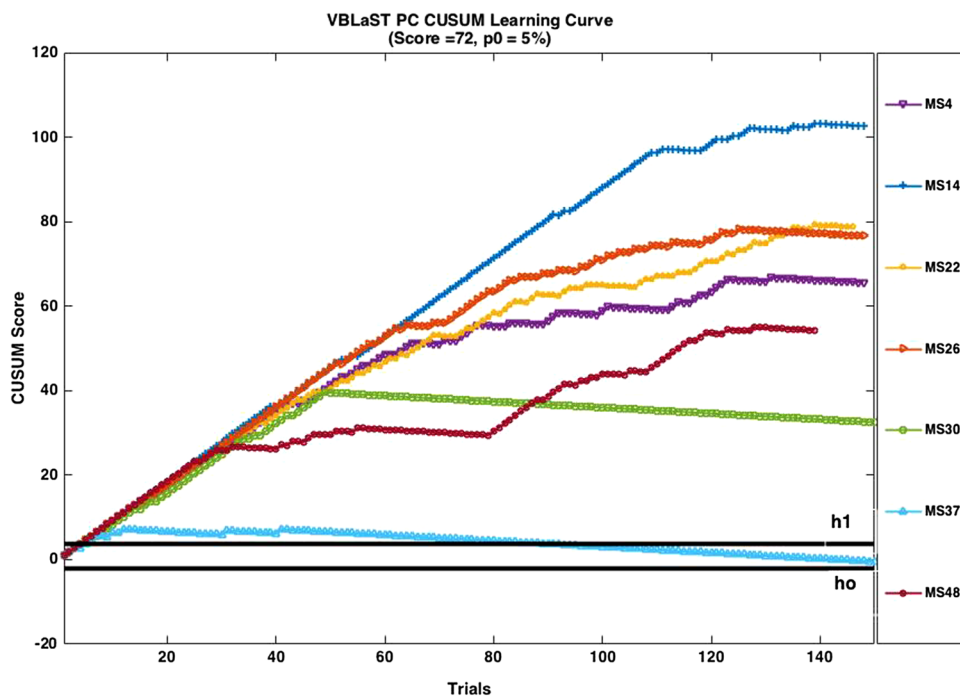
The inverse curve fitting results for the FLS-PC and the VBLaST-PC groups are shown in Fig. 8A, B. The learning plateau and the learning rate calculated from inverse curve fitting are presented in Table 5. The FLS group achieved a



**Fig. 6** CUSUM learning curves for medical students trained on VBLaST-PC simulator using the intermediate criterion success score of 56, acceptable failure rate  $p_0=5\%$



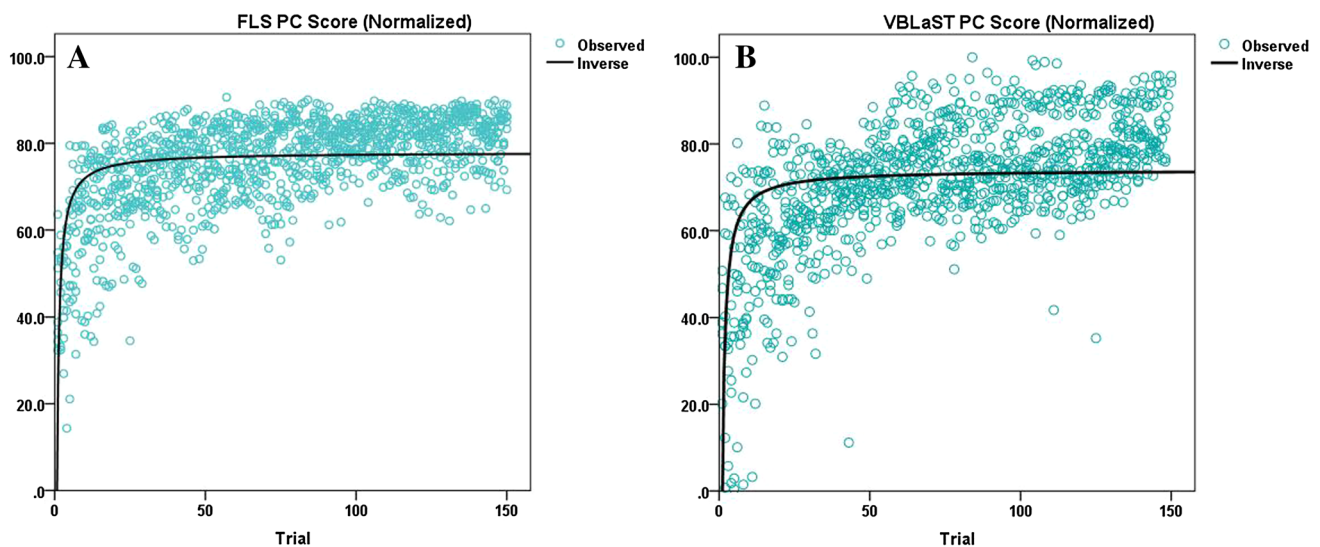
**Fig. 7** CUSUM learning curves for medical students trained on VBLaST-PC simulator using the FLS proficiency criterion success score of 72, acceptable failure rate  $p_0=5\%$



higher plateau (77.93,  $p < .001$ ) compared to the VBLaST group (70.00,  $p < .001$ ) in normalized scores. The FLS group also had a higher learning rate of seven trials compared to the VBLaST group which required an average of ten trials to reach 90% of the plateau.

**Discussion**

Learning curves capture gains in performance with repetition and have been widely used in assessing learning in surgical tasks [17–28]. There are multiple types of learning curves (linear, S-shaped, and positively or negatively



**Fig. 8** Inverse curve fitting for the derivation of learning plateau and learning rate based on **A** normalized FLS scores and **B** normalized VBLaST scores

accelerated) [29]. Although learning curves are useful, they are not adequate as it is hard to assess performance over time with knowledge of only the performance plateau and time to reach that level. Cumulative Summation Analysis (CUSUM) is a method where performance over time can be studied with set criteria [15]. CUSUM has been widely used in surgery to assess the learning progress for many different procedures [17, 18, 21, 23, 25, 27, 28, 30–51].

In our previous study on assessing learning of the VBLaST Peg Transfer task [17], we had analyzed the performance using three criteria (junior, intermediate, and senior) based on the classification reported by Fraser et al. [27]. In this work, we analyzed the learning of the VBLaST-PC task with a study using medical students.

Based on the two criteria (intermediate and proficient), four students in FLS and two students in VBLaST reached the acceptable failure rate of 5%. For the proficient score, none of the medical students were able to reach the acceptable failure rate at the end of 150 trials for both the simulators. It should be noted that all but one subject showed improvement in performance with a transition point on the CUSUM learning curve. Compared to peg transfer, the PC

task is more challenging and 150 trials were not enough to train subjects at the 5% failure rate.

When comparing the learning plateau reached by both groups on their respective simulators and learning rate, the FLS group achieved slightly higher learning plateau than VBLaST. The subjects in the FLS group also took three fewer trials than the VBLaST group to reach 90% of the plateau. Based on our experience working with VR simulators, we attribute this difference to the fact that subjects needed more time to become familiarized with the VR and haptics technology in the simulator. This hypothesis is difficult to test and may become self-evident in time as technology in VR and haptics becomes more sophisticated.

In terms of learning, subjects in the control condition did not show any learning on the FLS at the end of the study, with a total exposure of three trials on each of the simulators. They showed continual improvement on the VBLaST simulator from pretest to posttest to retention test, as indicated by the increasing test scores (Fig. 3C). This may suggest that the virtual simulator is more conducive to learning than the physical simulator during the trainee's initial exposure to simulation training. This finding may have important implications for surgical education when time availability for training is limited.

Compared to the control group, subjects who received training on the simulators improved their skills significantly after the 3 weeks of practice on their assigned simulator only. Based on the subjects' testing results (pretest–posttest), learning occurred from day 1 to day 15 of training. That is, VBLaST-trained subjects significantly improved in performance at posttest in the VBLaST simulator but not in the FLS simulator, whereas FLS-trained

**Table 5** Learning plateau and learning rate for subjects trained on FLS or VBLaST

Simulator	Measure	Learning plateau	Learning rate (trials)
FLS-PC	Score	77.93	7
VBLaST-PC	Score	74.00	10

subjects showed significant improvement in the FLS simulator but not in the VBLaST simulator. This is to be expected due to specificity in learning.

Trained subjects were able to retain their skills even after a period of non-use beyond the training period. In fact, subjects who were trained on the VBLaST continued to improve their performance on the FLS simulator at retention test. This may imply that skills learned on the VBLaST simulator are being transferred to the FLS environment. Similarly, the continual improvement shown by FLS-trained subjects on the VBLaST at pretest and post-test suggests that there is some transfer of learning from the FLS environment.

Overall, our study highlights that laparoscopic surgical skill training in a virtual environment is comparable to training in a physical environment, taking into account additional time needed for familiarization with the VR environment. This is a very important finding when it comes to planning simulation center experiences for surgical trainees. Repeated practice on the VBLaST simulator does not require expensive consumables and does not require replacing the materials between trials. The virtual environment is also capable of providing immediate feedback on time, error, and score, whereas the FLS simulator requires a proctor to keep time and calculate error and score, which is time consuming, labor intensive and results in significantly delayed feedback. The virtual environment is also capable of providing adaptive learning and one can imagine that a simulator can be programmed at different levels of difficulty with progression of training.

Future directions for this research include studies to investigate the transfer of learning from one simulator environment to the other and, by extension, the transfer of these laparoscopic surgery skills from the simulation lab to the operating theater. This will allow us to validate the predictive power of the simulators as a training tool for surgical skills mastery.

**Acknowledgements** The authors would like to thank Drs. Scott Epstein, Jesse Rideout, George Perides, and Christopher Awtrey for their assistance with the experimental setup, and Jannine Dewar, Winnie Chen, Yiman Lou, Emily Diller, Nicole Santos, and Jamaya Carter for their assistance in collecting and analyzing the performance data.

**Funding** This work was supported by a grant from the National Institutes of Health (NIBIB R01 EB010037-01).

**Compliance with ethical standards**

**Disclosures** Drs. Linsk, Monden, Sankaranarayanan, Ahn, De, and Cao have no conflicts of interest or financial ties to disclose. Dr. Jones is a Consultant for Allurion and The Medical Company. Dr. Schweitzberg is a Consultant for Covidien, and Great Venture Partner, and has equity interest in Human Extension, Acuity Bio, Arch Therapeutics, and Gordian Medical.

## References

1. Dawson SL, Kaufman JA (1998) The imperative for medical simulation. In: Proceedings of the IEEE, pp 479–483
2. Cao CGL, MacKenzie C, Payandeh S (1996) Task and motion analyses in endoscopic surgery. In: Proceedings ASME Dynamic Systems and Control Division (Fifth Annual Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems), pp 583–590
3. Fried GM (2008) FLS assessment of competency using simulated laparoscopic tasks. *J Gastrointest Surg* 12(2):210–212
4. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillan LF, Schirmer B et al (2004) Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 135(1):21–27. <http://www.ncbi.nlm.nih.gov/pubmed/14694297>
5. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM (2003) Evaluating laparoscopic skills, setting the pass/fail score for the MISTELS system. *Surg Endosc Other Interv Tech* 17(6):964–967
6. Arikatla VS, Sankaranarayanan G, Ahn W, Chellali A, Cao CGL, De S (2013) Development and validation of VBLaST-PT<sup>®</sup>: a virtual peg transfer simulator. In: Studies in health technology and informatics. IOS Press, Amsterdam, pp. 24–30. <http://www.ncbi.nlm.nih.gov/pubmed/23263645>
7. Sankaranarayanan G, Adair JD, Halic T, Gromski MA, Lu Z, Ahn W et al (2011) Validation of a novel laparoscopic adjustable gastric band simulator. *Surg Endosc* 25(4):1012–1018. <http://www.ncbi.nlm.nih.gov/pubmed/20734069>
8. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK et al (2002) Virtual reality training improves operating room performance. *Ann Surg* 236(4):458–464. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1422600/>
9. Seymour NE (2008) VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World J Surg* 32(2):182–188. <http://www.ncbi.nlm.nih.gov/pubmed/18060453>
10. Sankaranarayanan G, Lin H, Arikatla VS, Mulcare M, Zhang L, Derevianko A et al (2010) Preliminary face and construct validation study of a virtual basic laparoscopic skill trainer. *J Laparoendosc Adv Surg Tech* 20(2):153–157. <http://www.liebertonline.com/doi/abs/10.1089/lap.2009.0030>
11. Chellali A, Ahn W, Sankaranarayanan G, Flinn JT, Schweitzberg SD, Jones DB et al (2015) Preliminary evaluation of the pattern cutting and the ligating loop virtual laparoscopic trainers. *Surg Endosc* 29(4):815–821. <http://www.ncbi.nlm.nih.gov/pubmed/25159626>
12. Arikatla VS, Sankaranarayanan G, Ahn W, Chellali A, De S, Caroline GL et al (2013) Face and construct validation of a virtual peg transfer simulator. *Surg Endosc* 27(5):1721–1729. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3625247&tool=pmcentrez&rendertype=abstract>
13. Chellali A, Zhang L, Sankaranarayanan G, Arikatla VS, Ahn W, Derevianko A et al (2014) Validation of the VBLaST peg transfer task: a first step toward an alternate training standard. *Surg Endosc*. <http://www.ncbi.nlm.nih.gov/pubmed/24771197>
14. Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R (2007) Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care* 16(3):203–207. <http://qualitysafety.bmj.com/content/16/3/203>
15. Williams SM, Parry BR, Schlup MM (1992) Quality control: an application of the cusum. *Br Med J* 304(6838):1359–1361. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1882013/>
16. Feldman LS, Cao J, Andalib A, Fraser S, Fried GM (2009) A method to characterize the learning curve for performance of a

- fundamental laparoscopic simulator task: defining “learning plateau” and “learning rate”. *Surgery* 146(2):381–386. <http://www.ncbi.nlm.nih.gov/pubmed/19628099>
17. Zhang L, Sankaranarayanan G, Arikatla VS, Ahn W, Grosdemouge C, Rideout JM et al (2013) Characterizing the learning curve of the VBLaST-PT<sup>®</sup> (Virtual Basic Laparoscopic Skill Trainer). *Surg Endosc* 27(10):3603–3615. <http://www.ncbi.nlm.nih.gov/pubmed/23572217>
  18. Fraser SA, Bergman S, Garzon J (2012) Laparoscopic splenectomy: learning curve comparison between benign and malignant disease. *Surg Innov* 19(1):27–32. <http://www.ncbi.nlm.nih.gov/pubmed/21719436>
  19. Melfi FMA, Mussi A (2008) Robotically assisted lobectomy: learning curve and complications. *Thorac Surg Clin* 18(3):289–295, vi–vii. <http://www.sciencedirect.com/science/article/pii/S1547412708000492>
  20. Cook JA, Ramsay CR, Fayers P (2007) Using the literature to quantify the learning curve: a case study. *Int J Technol Assess Health Care*. 23(2):255–260. <http://www.ncbi.nlm.nih.gov/pubmed/17493312>
  21. Li JCM, Lo AWI, Hon SSF, Ng SSM, Lee JFY, Leung KL (2012) Institution learning curve of laparoscopic colectomy—a multi-dimensional analysis. *Int J Colorectal Dis* 27(4):527–533. <http://www.ncbi.nlm.nih.gov/pubmed/22124675>
  22. Chang L, Satava RM, Pellegrini CA, Sinanan MN (2003) Robotic surgery: identifying the learning curve through objective measurement of skill. *Surg Endosc* 17(11):1744–1748. <http://www.ncbi.nlm.nih.gov/pubmed/12958686>
  23. Jaffer U, Cameron AEP (2008) Laparoscopic appendectomy: a junior trainee’s learning curve. *J Soc Laparoendosc Surg* 12(3):288–291. <http://www.ncbi.nlm.nih.gov/pubmed/18765054>
  24. Veronesi G (2014) Robotic thoracic surgery technical considerations and learning curve for pulmonary resection. *Thorac Surg Clin* 24(2):135–141. <http://dx.doi.org/10.1016/j.thorsurg.2014.02.009>
  25. Bokhari MB, Patel CB, Ramos-Valadez DI, Ragupathi M, Haas EM (2011) Learning curve for robotic-assisted laparoscopic colorectal surgery. *Surg Endosc* 25(3):855–860. <http://www.ncbi.nlm.nih.gov/pubmed/20734081>
  26. Petersen RH, Hansen HJ (2012) Learning curve associated with VATS lobectomy. *Ann Cardiothorac Surg* 1(1):47–50
  27. Fraser SA, Feldman LS, Stanbridge D, Fried GM (2005) Characterizing the learning curve for a basic laparoscopic drill. *Surg Endosc* 19(12):1572–1578. <http://www.ncbi.nlm.nih.gov/pubmed/16235127>
  28. Son G-M, Kim J-G, Lee J-C, Suh Y-J, Cho H-M, Lee Y-S et al (2010) Multidimensional analysis of the learning curve for laparoscopic rectal cancer surgery. *J Laparoendosc Adv Surg Tech* 20(7):609–617. <http://www.ncbi.nlm.nih.gov/pubmed/20701545>
  29. Magill RA (2007) *Motor learning and control: concepts and applications*, 8th edn. McGraw-Hill, New York. <http://www.amazon.com/dp/0073047325>
  30. Kinsey SE, Giles FJ, Holton J (1989) Cusum plotting of temperature charts for assessing antimicrobial treatment in neutropenic patients. *BMJ* 299(6702):775–776
  31. Calsina L, Clará A, Vidal-Barraquer F (2011) The use of the CUSUM chart method for surveillance of learning effects and quality of care in endovascular procedures. *Eur J Vasc Endovasc Surg* 41(5):679–684. <http://www.ncbi.nlm.nih.gov/pubmed/21333562>
  32. Okrainec A, Ferri LE, Feldman LS, Fried GM (2011) Defining the learning curve in laparoscopic paraesophageal hernia repair: a CUSUM analysis. *Surg Endosc* 25(4):1083–1087. <http://www.ncbi.nlm.nih.gov/pubmed/20835725>
  33. Young A, Miller JP, Azarow K (2005) Establishing learning curves for surgical residents using Cumulative Summation (CUSUM) Analysis. *Curr Surg* 62(3):330–334. <http://www.ncbi.nlm.nih.gov/pubmed/15890218>
  34. Van Rij AM, McDonald JR, Pettigrew RA, Putterill MJ, Reddy CK, Wright JJ (1995) Cusum as an aid to early assessment of the surgical trainee. *Br J Surg* 82(11):1500–1503. <http://www.ncbi.nlm.nih.gov/pubmed/8535803>
  35. Naik VN, Devito I, Halpern SH (2003) Cusum analysis is a useful tool to assess resident proficiency at insertion of labour epidurals. *Can J Anaesth* 50(7):694–698. <http://www.ncbi.nlm.nih.gov/pubmed/12944444>
  36. Bartlett A, Parry B (2001) Cusum analysis of trends in operative selection and conversion rates for laparoscopic cholecystectomy. *ANZ J Surg* 71(8):453–456. <http://www.ncbi.nlm.nih.gov/pubmed/11504287>
  37. McCarter FD, Luchette FA, Molloy M, Hurst JM, Davis K, Johannigman JA et al (2000) Institutional and individual learning curves for focused abdominal ultrasound for trauma. *Ann Surg* 231(5):689–700. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1421056/>
  38. Papanra R, Biau DJ, Mann LK, Johnson A, Moise KJ Jr (2011) Use of the Learning Curve–Cumulative Summation test for quantitative and individualized assessment of competency of a surgical procedure in obstetrics and gynecology: fetoscopic laser ablation as a model. *Am J Obstet Gynecol* 204(3):218.e1–218.e9. <http://www.sciencedirect.com/science/article/pii/S0002937810022258>
  39. Zhang L, Sankaranarayanan G, Arikatla VS, Ahn W, Grosdemouge C, Rideout JM et al (2013) Characterizing the learning curve of the VBLaST-PT<sup>®</sup> (Virtual Basic Laparoscopic Skill Trainer). *Surg Endosc*. <http://www.ncbi.nlm.nih.gov/pubmed/23572217>
  40. Lerch L, Donald JC, Olivetto IA, Lesperance M, van der Westhuizen N, Rusnak C et al (2007) Measuring surgeon performance of sentinel lymph node biopsy in breast cancer treatment by cumulative sum analysis. *Am J Surg* 193(5):556–560. <http://www.sciencedirect.com/science/article/pii/S0002961007000918>
  41. Tekkis PP, Senagore AJ, Delaney CP, Fazio VW (2005) Evaluation of the learning curve in laparoscopic colorectal surgery. *Ann Surg* 242(1):83–91. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1357708/>
  42. Pendlimari R, Holubar SD, Dozois EJ, Larson DW, Pemberton JH, Cima RR (2012) Technical proficiency in hand-assisted laparoscopic colon and rectal surgery: determining how many cases are required to achieve mastery. *Arch Surg* 147(4):317–322. <http://www.ncbi.nlm.nih.gov/pubmed/22184135>
  43. Forbes TL, DeRose G, Kribs SW, Harris KA (2004) Cumulative sum failure analysis of the learning curve with endovascular abdominal aortic aneurysm repair. *J Vasc Surg* 39(1):102–108. <http://www.sciencedirect.com/science/article/pii/S0741521403009224>
  44. Novotný T, Dvůrák M, Staffa R (2011) The learning curve of robot-assisted laparoscopic aortofemoral bypass grafting for aortoiliac occlusive disease. *J Vasc Surg* 53(2):414–420. <http://www.ncbi.nlm.nih.gov/pubmed/21093201>
  45. Buchs NC, Pugin F, Bucher P, Hagen ME, Chassot G, Koutny-Fong P et al (2012) Learning curve for robot-assisted Roux-en-Y gastric bypass. *Surg Endosc* 26(4):1116–1121. <http://www.ncbi.nlm.nih.gov/pubmed/22044973>
  46. Kye B-H, Kim J-G, Cho H-M, Kim H-J, Suh Y-J, Chun C-S (2011) Learning curves in laparoscopic right-sided colon cancer surgery: a comparison of first-generation colorectal surgeon to advance laparoscopically trained surgeon. *J Laparoendosc Adv Surg Tech* 21(9):789–796. <http://www.ncbi.nlm.nih.gov/pubmed/21854205>
  47. Filho DO, Rodrigues G (2002) The construction of learning curves for basic skills in anesthetic procedures: an application for the cumulative sum method. *Anesth Analg* 95(2):411–416. <http://www.anesthesia-analgesia.org/content/95/2/411>

48. Murzi M, Cerillo AG, Bevilacqua S, Gilmanov D, Farneti P, Glauber M (2012) Traversing the learning curve in minimally invasive heart valve surgery: a cumulative analysis of an individual surgeon's experience with a right minithoracotomy approach for aortic valve replacement. *Eur J Cardiothorac Surg* 41(6):1242–1246. <http://www.ncbi.nlm.nih.gov/pubmed/22232493>
49. Li X, Wang J, Ferguson MK (2014) Competence versus mastery: the time course for developing proficiency in video-assisted thoracoscopic lobectomy. *J Thorac Cardiovasc Surg* 147(4):1150–1154. [10.1016/j.jtcvs.2013.11.036](https://doi.org/10.1016/j.jtcvs.2013.11.036)
50. Schlup MMT, Williams SM, Barbezat GO (1997) ERCP: a review of technical competency and workload in a small unit. *Gastrointest Endosc* 46(1):48–52. <http://www.sciencedirect.com/science/article/pii/S0016510797702098>
51. Bould MD, Crabtree NA, Naik VN (2009) Assessment of procedural skills in anaesthesia. *Br J Anaesth* 103(4):472–483. <http://bjaoxfordjournals.org/content/103/4/472>