Published in final edited form as:

Surg Endosc. 2013 October ; 27(10): 3603–3615. doi:10.1007/s00464-013-2932-5.

Characterizing the Learning Curve of the VBLaST-PT® (Virtual Basic Laparoscopic Skill Trainer)

Likun Zhang, MS¹, Ganesh Sankaranarayanan, PhD², Venkata Sreekanth Arikatla, MS², Woojin Ahn, PhD², Cristol Grosdemouge, MS¹, Jesse M. Rideout, MD³, Scott K. Epstein, MD³, Suvranu De, ScD², Steven D. Schwartzberg, MD⁵, Daniel B. Jones, MD⁴, and Caroline G.L. Cao, PhD^{1,5,6}

¹Department of Mechanical Engineering, Tufts University

²Center for Modeling, Simulation and Imaging in Medicine, Rensselaer Polytechnic Institute

³School of Medicine, Tufts University

⁴Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School

⁵Department of Surgery, Cambridge Health Alliance, Harvard Medical School

⁶Department of Biomedical, Industrial and Human Factors Engineering, Wright State University

Abstract

Background and study aim—Mastering laparoscopic surgical skills requires considerable time and effort. The Virtual Basic Laparoscopic Skill Trainer (VBLaST-PT®) is being developed as a computerized version of the peg transfer task of the Fundamentals of Laparoscopic Surgery (FLS) system using virtual reality technology. We assessed the learning curve of trainees on the VBLaST-PT® using the cumulative summation (CUSUM) method and compared them with those on the FLS to establish convergent validity for the VBLaST-PT®.

Methods—Eighteen medical students from were assigned randomly to one of three groups: control, VBLaST-training and FLS-training. The VBLaST and the FLS groups performed a total of 150 trials of the peg-transfer task over a three week period, five days a week. Their CUSUM scores were computed based on pre-defined performance criteria (junior, intermediate and senior levels).

Results—Of the six subjects in the VBLaST-training group, five achieved at least the “junior” level, three achieved the “intermediate” level and one achieved the “senior” level of performance criterion by the end of the 150 trials. In comparison, for the FLS group, three students achieved the “senior” criterion and all six students achieved the “intermediate” and “junior” criteria by the 150th trials. Both the VBLaST-PT® and the FLS systems showed significant skill improvement and retention, albeit with system specificity as measured by transfer of learning in the retention test: The VBLaST-trained group performed better on the VBLaST-PT® than on FLS ($p=0.003$), while the FLS-trained group performed better on the FLS than on VBLaST-PT® ($p=0.002$).

Corresponding Author: Caroline G.L. Cao, PhD, Professor of Biomedical, Industrial and Human Factors Engineering, Wright State University, Dayton, OH 45435, Ph: (937) 775-5044, caroline.cao@wright.edu.

Disclosure

Likun Zhang, Venkata Sreekanth Arikatla, Cristol Grosdemouge, and Drs. Ganesh Sankaranarayanan, Woojin Ahn, Jesse M. Rideout, Scott K. Epstein, Daniel B. Jones, Suvranu De and Caroline G.L. Cao have no conflicts of interest or financial ties to disclose. Steven D. Schwartzberg had a research grant from Ethicon and is a consultant for Olympus, Styker, MITI, Acuity Bio, Neatstich and Surgicquest. Daniel B. Jones is a consultant for Allurion.

Conclusion—We characterized the learning curve for a virtual peg transfer task on the VBLaST-PT[®] and compared it with the FLS using CUSUM analysis. Subjects in both training groups showed significant improvement in skill performance, but the transfer of training between systems was not significant.

Keywords

Learning curve; Cumulative summation; CUSUM; virtual reality; surgical training; convergent validity

Introduction

Performing laparoscopic procedures requires a considerable amount of hand-eye coordination skill due to the limited two-dimensional visual feedback while operating in a three-dimensional environment with long slender tools [1]. The movement of the tools is also counterintuitive due to the “fulcrum effect” in which the surgeons’ hand movements and the tool tip movements are reversed [2]. These psychomotor challenges are overcome only by extensive practice, which is required to achieve proficiency in laparoscopic skills [3]. The importance of skill mastery in laparoscopic surgery to patient safety has prompted many institutions to implement a targeted training curriculum to train surgical residents, using simulation technology [4].

The McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) is a validated system that was developed to teach and measure basic laparoscopic skills [5]. This system has been adopted by the Fundamentals of Laparoscopic Surgery (FLS) committee of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) and the program is administered in partnership with the American College of Surgeons (ACS) as the standard to assess the proficiency of laparoscopic skills [6, 7]. Since 2009, successful completion of the FLS exam is a requirement before being allowed to take the Qualifying Examination of the American Board of Surgery. Major drawbacks to the FLS practical exam component include the difficulty to evaluate performance objectively and the time needed to score manually. The Virtual Basic Laparoscopic Skill Trainer (VBLaST[®]) [8] is a virtual reality simulator that is being developed to simulate the FLS tasks with funding support from the National Institutes of Health (NIH). The VBLaST-PT[®] is the version that simulates the peg transfer task of the FLS. Unlike other trainers, the VBLaST-PT[®] system is capable of measuring performance without the need for a proctor or replenishing training and testing materials, two significant cost drivers in the FLS program.

It is imperative for any training program for laparoscopic skill acquisition to objectively analyze the trainees’ learning outcomes, monitoring their progress continuously to make sure that the required skills are being learned, and providing feedback to remedy any shortcomings. This can be achieved by examining the characteristics of the learning curve.

The cumulative summation (CUSUM) is a criterion-based method that is widely 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 [9, 10].

The objective of this study was to show convergent validity of the VBLaST-PT[®] system as a laparoscopic training simulator. 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 hypothesized that the learning curves on the VBLaST-PT[®] and FLS would be similar, with performance improving with practice. In addition, subjects with training on either simulator would perform better than those with no training in the post and retention tests. Learning curve characteristics of the peg transfer task on both simulation systems were assessed using the CUSUM method.

Methodology

Subjects

Eighteen medical students (six per group) at a local medical school responded to a recruiting email. They fit the selection criteria of having little or no prior experience with surgery or surgical simulators, normal or corrected to normal vision, and no motor impairment that prevented manipulation of two laparoscopic tools in the surgical simulators. Subjects were compensated for their participation.

Simulators

The FLS system used was the standard SAGES-approved trainer box with a membrane cover and two slits to place 12-mm trocars, as shown in Figure 1a. The peg board for the task was placed inside the box by securing it to the Velcro clip. A camera with fixed focal length captured the view of the task space and displayed it on a monitor for the subject. A digital video capture device (AVerMedia, Milpitas, CA, USA) was used to record subjects' performance inside the task space. The video was used to extract timing and error measurements for data analysis.

The VBLaST-PT[®] system consisted of two laparoscopic tools connected to haptic devices (Figure 1b) mounted in front of a monitor, and a virtual reality environment simulating the FLS peg transfer task. The computational software in the VBLaST-PT[®] system uses rigid body dynamics to simulate the interaction between the tools, and objects in the virtual environment. The two PHANTOM Omni haptic devices (SensAble Technologies, Wilmington, MA, USA) connected to the instrumented tools provide force feedback to the user. 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.

Experiment Design and Procedure

The peg transfer task, the first of the five tasks in the FLS, was used in this convergent validation study. In this task, a set of six pegs were transferred from one side of the board to the other and back by picking each peg with one tool (operated by the dominant hand to start) and transferring it to the other tool before placing it on the board. The task was designed to develop coordination skill in both hands for performing laparoscopic surgery. A maximum of 300 seconds is allotted for subjects to complete the task.

The experiment was a between-subjects design with three different training conditions. Subjects were randomly assigned to one of the three conditions: control, VBLaST and FLS. Subjects in the control group did not receive any training on the task, while those in the other two training groups received training on the assigned simulator over a period of three weeks.

In this Institutional Review Board (IRB) approved study, before the testing session began, all subjects watched an instructional video that demonstrated the proper procedure to perform the peg transfer task in both the FLS and VBLaST-PT[®] systems. All subjects then performed the task once using both simulators to mark their own baseline performance (also

pre-test). The order of simulators was counterbalanced. That is, half of the subjects used the FLS first for the pre-test, while the other half of the subjects used the VBLaST simulator first in the pre-test.

Subjects assigned to a training condition then practiced the task on the assigned simulator for a session each day, five days a week, for three consecutive weeks for a total of 15 training sessions. During the training sessions, subjects performed 10 trials of the task or attempted as many trials as possible during half an hour, whichever was shorter. The VBLaST-PT[®] system recorded subject performance at the end of each trial automatically, whereas the FLS performance scores were computed manually using the video recording of the sessions. Those subjects in the control condition did not practice the task at all.

At the end of the three weeks, all subjects performed the peg transfer task to mark their final performance (post-test) using both simulators. The order of simulators was counterbalanced. Two weeks following the post-test, all subjects performed the task on both simulators again to determine their skill retention.

Dependent measures

Task completion time and errors (dropped pegs outside of board) were the dependent measures. Performance scores on each simulator were calculated based on the established scoring metrics for the FLS tasks [5]. The scores for each training group were then normalized separately. For the FLS condition, subjects' scores were normalized to the best published score in the literature [12], whereas the VBLaST subjects' scores were normalized to the best expert score thus far available on the VBLaST-PT[®] in the ongoing multi-institution concurrent validation study [11].

CUSUM calculation

CUSUM charts were generated for each individual subject, and for each training group as a whole. Three performance criteria representing 'junior', 'intermediate' and 'senior' level of performance were defined for the individual and overall learning curves. The FLS learning curve performance criteria were derived from the MISTELS database [12] (mean junior criterion score = 41, mean intermediate criterion score = 65, and mean senior criterion score = 76). The VBLaST-PT[®] criteria were drawn from the ongoing multi-institution concurrent validation study (mean junior criterion score = 59, mean intermediate criterion score = 64, and mean senior criterion score = 68) [11]. For each of the criterion scores, the failure rates were calculated. When the computed peg transfer score for each trial is equal to or better than the criterion score, it was coded as a 'success' (1), whereas a lower score was coded as a 'failure' (0) (Table 2). The acceptable failure rate (p_0) was set at 5% for each criterion level, while the unacceptable failure rate (p_1) was set at 10% ($2 \times p_0$) [12]. Type I and type II errors (α and β) were set at 0.05 and 0.20, respectively. Based on these parameters, two decision limits (h_0 and h_1) and the s value for each successive trial were calculated in a running tally. 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 plotted CUSUM line indicates success, whereas a positive slope suggests failure. This procedure was performed for each subject using all three criterion scores of performance. Table 1 shows the CUSUM variables for data analysis. Table 2 is an example of the CUSUM chart for a representative subject (medical student no. 8) using the VBLaST-PT[®] intermediate criterion score of 64.

Data Analysis

For the statistical analysis of pre, post and retention tests, the SPSS 18.0 statistical software package (IBM Inc.) was used. Separate two-way mixed factorial ANOVAs with one within-

subject variable (pre, post and retention test scores) and one between-subject variable (training group) were used for the FLS and the VBLaST-PT[®] scores. ANOVA results with significance ($p < 0.05$) were further analyzed (post-hoc comparison) using the Tukey HSD method. Paired-sample t-tests with a Bonferroni correction ($p < 0.002$ for significance) was used to further compare the performance between pre, post and retention tests for each of the simulators.

Results

Individual CUSUM learning curve for VBLaST

Based on the VBLaST-PT[®] junior criterion (score = 59), five out of six medical students (MS) achieved the acceptable 5% failure rate the 150th peg transfer trial (MS 8 achieved it at the 73rd trial, MS 9 and 11 at the 145th trial, MS 10 at the 102nd trial, MS 12 at the 59th trial; range = 59 to 145 trials) (see Figure 2). Individual medical student failure rates ranged from 1.3% to 10.1% for this criterion. Note that only MS 8, MS 10 and MS 12 finished all 150 trials. All other subjects in this group could not finish 10 trials within half an hour during the first training session. Using the VBLaST-PT[®] junior criterion, all subjects trained on VBLaST-PT[®] demonstrated a transition point at which their CUSUM line changed from a positive to a negative slope, suggesting a trend toward more successful performance (MS 7 at trial 19, MS 8 at trial 3, MS 9 at trial 16, MS 10 at trial 4, MS 11 at trial 11, MS 12 at trial 6; transition point range = 3 to 19 trials).

Based on the VBLaST-PT[®] intermediate criterion (score = 64), only three medical students trained on VBLaST-PT[®] achieved an acceptable 5% failure rate by the 150th peg transfer trial (MS 8 and 12 at trial 73, MS 10 at trial 130) (see Figure 4). Individual student failure rates ranged from 2.0% to 12.8%. Using the VBLaST-PT[®] intermediate criterion, all students trained on VBLaST-PT[®] had a performance transition point (MS 7 at trial 19, MS 8 at trial 3, MS 9 at trial 16, MS 10 at trial 4, MS 11 at trial 19, MS 12 at trial 6; transition point range = 3 to 19 trials).

Using the VBLaST-PT[®] senior criterion (score = 68), only two medical students trained on VBLaST-PT[®] achieved the acceptable 5% failure rate by the 150th peg transfer trial (MS 8 at trial 102, MS 12 at trial 73) (see Figure 6). Individual student failure rates ranged from 2.0% to 12.8%. Under the VBLaST-PT[®] senior criterion, all students trained on VBLaST-PT[®] had a performance transition point (MS 7 at trial 19, MS 8 at trial 7, MS 9 at trial 16, MS 10 at trial 21, MS 11 at trial 22, MS 12 at trial 6; transition point range = 6 to 22 trials).

Individual CUSUM learning curve for FLS

Based on the FLS junior criterion (score = 41), all medical students trained on FLS achieved the acceptable 5% failure rate by the 150th peg transfer trial (MS 2 at trial 59, MS 1, 3, 4, 5, 6 at trial 30) (see Figure 3). MS 2 had a failure rate of 1.3%, while all others in the group had a failure rate of 0. Under the FLS junior criterion, all students trained on FLS demonstrated a transition point at which their CUSUM line changed from a positive to a negative slope, suggesting a trend toward more successful trials (MS 2 at trial 2; MS 1, 3-6 at trial 1).

Using the FLS intermediate criterion (score = 65), all medical students trained on FLS achieved a 5% failure rate by the 150th peg transfer trial (MS 1 at trial 45, MS 2 at trial 116, MS 3 at trial 87, MS 4 at trial 102, MS 5 at trial 59, MS 6 at trial 30; range = 30 to 116 trials) (see Figure 5). Individual student failure rates ranged from 0 to 4%. Under the FLS intermediate criterion, all students trained on FLS had a performance transition point (MS 1 at trial 5, MS 2 at trial 36, MS 3 at trial 5, MS 4 at trial 11, MS 5 and MS 6 at trial 1; transition point range = 1 to 36 trials).

Under the FLS senior criterion (score = 76), only three medical students trained on FLS achieved a 5% failure rate by the 150th peg transfer trial (MS 1 at trial 116, MS 3 at trial 145, MS 6 at trial 59) (see Figure 7). Individual student failure rates ranged from 1.3% to 10.7%. Based on the FLS senior criterion, all students trained on FLS had a performance transition point (MS 1 at trial 9, MS 2 at trial 50, MS 3 at trial 8, MS 4 at trial 41, MS 5 at trial 59, MS 6 at trial 27; transition point range = 8 to 59 trials).

The results from the CUSUM individual learning curve analysis are summarized in Table 3.

Total CUSUM learning curve

Based on the junior criterion, the acceptable failure rate of 5% was achieved by trial 87 for the VBLaST group overall, and by trial 30 for the FLS group overall. The junior criterion transition point for the VBLaST group was at trial 4, whereas for the FLS group was at trial 1 (see Figure 8). The total junior criterion failure rate was 2.7% for the VBLaST-PT[®] and 0 for the FLS.

Under the intermediate criterion, the failure rate of 5% was achieved by trial 102 for the VBLaST group overall, with a transition point at trial 6. For the FLS group, the acceptable failure rate of 5% was achieved by trial 59, with a transition point at trial 2 (see Figure 9). Total intermediate criterion failure rate was 3.3% for the VBLaST-PT[®] group, but 1.3% for the FLS.

Using the senior criterion, the VBLaST group achieved a 5% failure rate by trial 145, with a transition point at trial 21, while the FLS group achieved a 5% failure rate by trial 102, with a transition point at trial 5 (see Figure 10). Total senior criterion failure rate was 5.3% for the VBLaST-PT[®] and 3.3% for the FLS.

Pre, post and retention test results

The normalized performance scores for the three training groups on VBLaST-PT[®] and FLS simulators are shown in Figures 11 and 12, respectively. Mixed factorial two-way (training group and test) ANOVA on the performance results from the VBLaST-PT[®] showed a significant main effect on the test factor ($p < 0.001$). The between-subjects factor of training also showed a significant effect ($p < 0.001$). Post-hoc analysis using the Tukey HSD method showed that the performance of the VBLaST trained group on the VBLaST-PT[®] during post-test and retention test was significantly better than both the FLS-trained ($p = 0.002$) and the control groups ($p < 0.001$). No significant difference was found between the performances of the FLS and the control group ($p = 0.375$). Paired-samples t-test analysis on the VBLaST-PT[®] scores showed a significant difference between the pre and post tests ($p < 0.002$) but no difference between the post and retention tests ($p = 0.159$), indicating skill retention.

On the FLS simulator, a significant main effect in the test factor ($p < 0.001$) was observed. The between-subjects factor of training also showed a significant effect ($p = 0.001$). Post-hoc analysis using the Tukey HSD method showed that the performance of the FLS trained group on the FLS simulator was significantly better than both the VBLaST trained ($p = 0.004$) and the control groups ($p = 0.001$). No significant difference was found between the performances of the VBLaST and the control group ($p = 0.575$). Paired-samples t-test showed a significant difference between pre and post-test ($p = 0.001$) but no significant difference between post and retention tests on the FLS for the FLS training group ($p = 0.486$), indicating skill retention.

Overall, both the VBLaST and the FLS trained groups performed better when tested on their respective simulators indicating skill acquisition. However, performance was not better than

the control group when tested on a different simulator than the one used for training, indicating no significant skill transfer between the VBLaST-PT[®] and the FLS systems.

Discussion

The performance of many motor skills improves with practice and repetition over time. A learning curve is used to visually describe the learning process. It can be linear, positively or negatively accelerated, or ogive or S-shaped [13]. In tasks such as surgery, the learning curve tends to be negatively accelerated with markedly increasing improvement early on and subsequent smaller improvements as the curve reaches a plateau. Cook et al. [14] described three main features of the learning curve: the “starting level” where the performance begins, “learning level” where the rate of learning determines how quickly a performance level is reached, and the “expert level” where the performance has reached the plateau. This characteristic curve can be observed in only 10 trials of practice for experts, and five hours of practice in beginners. Improvement in performance with just 10 trials has been shown for a group of 150 surgeons performing three basic tasks (rope pass drill, cup drop drill and triangle transfer drill) [15]. Similarly, other studies have shown that ten surgeons performing five tasks aimed at improving suturing and tying skills showed marked improvement in performance in four out of five tasks with just five hours of training [16]. In another study [17] that consisted of PGY2 and PGY3 residents, performing four tasks designed to test the dominant and non dominant hand and two-handed skills showed improved performance with only 4 to 5 hours of training. In our study, we have shown that our subjects reached their performance plateau before the 150th trial, or 7.5 hours of practice. Merely looking at the time to reach plateau, or the plateau level is not enough to assess the change of performance with time and practise over the course of the training program.

McCarter et al. [18] stated that the advantage of CUSUM analysis is that it allows the study of performance over time with experience as a continuous variable. Moreover, with CUSUM analysis a series of failures will be apparent by the change in the slope of the graph whereas random failures will not be detected since they do not affect the performance significantly. In their study on institutional and individual learning curves on focused abdominal ultrasound for trauma, they found that while the CUSUM analysis of the aggregate experience showed accuracy greater than 90%, individual experience showed a range of accuracy from 87% to 98%. The CUSUM can be used for monitoring progress of an individual while learning a new procedure, as well as to track the progress after the acquisition of the skill. In a retrospective analysis of colonoscopic examinations using the CUSUM technique for three different surgeons of various experience levels, it has been shown that an acceptable completion rate should be set above 90% [10]. In another retrospective study on Endoscopic Retrograde Cholangiopancreatography (ERCPs) performed over an 8 year-period by a single surgeon at Dunedin hospital using CUSUM analysis, satisfactory outcomes for selective cannulation were obtained after about 100 to 120 procedures [19]. Using CUSUM analysis it was shown that when ERCP was performed regularly, even small hospitals can develop and maintain expertise in this procedure. In a study on the performance of the surgical trainee using global assessment derived from modified Royal Australian College of Surgeon’s mentor form and CUSUM [20], after only 25 procedures, CUSUM was able to reliably report whether the trainee’s performance was satisfactory or not based on the criteria of the procedure duration.

The advantage of CUSUM is that it is an objective tool whereas global assessment rating is subjective in nature. In another study of the learning curve for laparoscopic colorectal resection of a colorectal unit using CUSUM with both the combined and the seven individual surgeon’s performance, steady state was reached after 310 cases with respect to conversion rate and at around 50 cases when considering intra-operative and major post-

operative complications [21]. More importantly the study showed that the continuous training of a new trainee did not affect the outcome of an established unit. In a retrospective study of technical proficiency of two surgeons on hand-assisted laparoscopic colon and rectal surgery using CUSUM analysis on the operative time, the change point from “learning period” to “skilled period” occurred after 108 and 105 cases for both the surgeons, respectively [22]. The “skilled period” operative time and complications were much lower compared to the “learning period”. The learning usually extends beyond fellowship years as was shown in this study using CUSUM analysis. Similar learning curve studies have been performed for several laparoscopic procedures [23–28], cardiac procedures [29], robotic surgeries [30–32], endovascular procedures [33], anesthesia [34] and many other procedures as well [18, 35–37].

Fraser et al. [12], used CUSUM to characterize the learning curve for the peg transfer task based on junior, intermediate and senior criterion standards. They asserted that even though supervisors can easily spot the general traits of a novice or an expert, less obvious traits may not be recognized by them. An objective, criteria-based method (CUSUM) has the advantage of being able to monitor trainees individually, tracking their progress based on specific set criteria. Based on their established junior, intermediate and senior criteria, by the 40th trial, one student achieved the same error rate (5%) as the senior level, three students achieved the error rate of intermediate, and ten students achieved the error rate of junior level. All the learning curves of the students showed transition points indicating the trend towards more successful trials.

In this work, the learning curve for both the FLS (derived from MISTELS) and VBLaST-PT[®] systems were established. For those six students who trained on the FLS, three students achieved a 5% failure rate for the senior criteria by the 150th trial, while all six students achieved the intermediate and junior criteria performance level. The key differences from the learning curve study on MISTELS [12] and our current study is the sample size (sixteen in [12] as opposed to six in this study) for the FLS training group and the number of trials (40 in [12] as opposed to 150 in this study). Nevertheless, our results showed similar or better performance on FLS with training with all the participants showing transition points in their CUSUM curve.

For the VBLaST training group, two students achieved the 5% failure rate needed for senior criteria, three students for the intermediate, and five students for the junior level by the 150th trial. Overall, the VBLaST-PT[®] system required more trials than FLS to achieve the three set criteria, as shown in the individual and group curves. Nevertheless, all students showed transition points in their CUSUM curve demonstrating the learning of skill in VBLaST-PT[®] system.

It is worth noting that even though the group CUSUM learning curves for the FLS and the VBLaST-PT[®] systems in our study showed that the training groups achieved the 5% failure rate, this does not imply that every single individual has achieved the 5% failure rate in either FLS or VBLaST-PT[®].

When the trainee is in the “learning level” of the learning curve [14], it must be ensured that the training does not stop there since fewer trials would have a greater effect on the learning and skill retention [38]. Scott and Jones [39] showed significant improvement in skills between second year medical students (MS2) and PGY3 in a learning curve study on five video-trainer tasks with 10 days of 30min training. In another study, Derossis et al. [40] showed that for two groups of surgical residents, the group that received training for five weeks on MISTELS tasks showed significant differences in scores compared to the group that did not receive any training at all. In another study by Kolozsvari et al. [41], whether

training and over-training in peg transfer task had any influence in learning intracorporeal suturing task was studied. They found that for novices, training in peg transfer was slightly associated with improvements in the learning curve for the intracorporeal suturing, and over-training in the peg transfer did not correlate with improved skill retention.

In our study, from the pre, post and retention tests, both the VBLaST and the FLS groups exhibited significant improvement in post-test scores and retention scores. In addition, the performance of the VBLaST and the FLS trained groups were best when tested on their respective simulators, showing specificity of training environments. When the VBLaST trained group was tested in the FLS, their performance was not significantly better than the control group. The FLS trained group showed the same specificity of training environment, and the same lack of transfer of learning compared to the control group.

Limitations and Future Work

Even though the VBLaST[®] system was designed to replicate the FLS tasks, it is nevertheless a virtual reality system as opposed to a real physical system. While this provides important benefits such as objective scoring, elimination of proctors, and unlimited self-replenishing materials, it still requires some adaptation by the users, especially when using it for the first time. To bridge this gap, we have since identified several issues that might contribute to the differences in the performance and their solutions have been implemented in the VBLaST-PT[®] simulator that is currently undergoing concurrent validation. For example, the workspace of the VBLaST-PT[®] simulator has since been modified to more closely match the FLS system so that the range of tool motion is more consistent with the movements in the FLS box. In addition, the tool and peg interaction have been modified so that the pegs can be picked using a grasper similar to the FLS system. Display enhancements such as shadows to provide clues about depth have also been implemented to enhance the performance in the VBLaST-PT[®] system.

The CUSUM method, though advantageous to monitor the progress of trainees using arbitrarily set criteria, cannot provide information on where the learning curve plateaus or the rate at which learning is achieved. Feldman et al. [42] used nonlinear regression to fit an inverse curve to estimate “learning plateau” and “learning rate” from the data collected in the study by Fraser et al. [12]. We plan to perform similar analyses in the data collected in this study to provide additional information on the learning curve of the individual students from the training groups.

In conclusion, the learning curve of the VBLaST-PT[®] simulator for the peg transfer task was characterized using the CUSUM method and compared with the currently established gold standard for training, the FLS system for the same task. We showed that all students trained on VBLaST-PT[®] showed transition points on their CUSUM curve indicating learning. The pre, post and retention tests also showed skill improvement and retention with training on VBLaST-PT[®] similar to FLS. However, the training effects on FLS and VBLaST are system-specific and do not appear to be transferrable to the other system. More work is needed to examine the issue of transfer of learning from VBLaST to the operating room.

Acknowledgments

Funding: NIBIB/NIH grant # R01EB010037

This work was supported by the NIBIB/NIH grant # R01EB010037. The authors thank the dedicated medical students who participated in the study. We also thank Saurabh Dargar for helping us to put together the VBLaST-PT[®] interface.

References

1. Shah J, Buckley D, Frisby J, Darzi A. Depth cue reliance in surgeons and medical students. *Surgical Endoscopy*. 2003; 17(9):1472–4. 10.1007/s00464-002-9178-y [PubMed: 12802650]
2. Gallagher AG, McClure N, McGuigan J, Ritchie K, Sheehy NP. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy*. 1998; 30(7):617–20. [PubMed: 9826140]
3. Adams JA. Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin*. 1987; 101(1):41–74.
4. Powers TW, Murayama KM, Toyama M, Murphy S, Denham EW 3rd, Derossis AM, et al. Housestaff performance is improved by participation in a laparoscopic skills curriculum. *American journal of surgery*. 2002; 184(6):626–9. discussion 9–30–9; discussion 9–30. [PubMed: 12488194]
5. Fraser SA, Klassen DR, Feldman LS, Ghitulescu GA, Stanbridge D, Fried GM. Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. *Surgical endoscopy*. 2003; 17(6):964–7. [PubMed: 12658417]
6. Fried GM. FLS assessment of competency using simulated laparoscopic tasks. *Journal of gastrointestinal surgery: official journal of the Society for Surgery of the Alimentary Tract*. 2008; 12(2):210–2. [PubMed: 17955315]
7. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004; 135(1):21–7. [PubMed: 14694297]
8. Sankaranarayanan G, Lin H, Arikatla VS, Mulcare M, Zhang L, Derevianko A, et al. Preliminary Face and Construct Validation Study of a Virtual Basic Laparoscopic Skill Trainer. *Journal of Laparoendoscopic & Advanced Surgical Techniques*. 2010; 20(2):153–7. [PubMed: 20201683]
9. Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RmS, Porcher Rl. Quality Control of Surgical and Interventional Procedures: A Review of the CUSUM. *Quality and Safety in Health Care*. 2007; 16(3):203–7. [PubMed: 17545347]
10. Williams SM, Parry BR, Schlup MM. Quality control: an application of the cusum. *BMJ : British Medical Journal*. 1992; 304(6838):1359–61.
11. Zhang L, Grosdemouge C, Arikatla VS, Ahn W, Sankaranarayanan G, De S, et al. The added value of virtual reality technology and force feedback for surgical training simulators. *Work: A Journal of Prevention, Assessment and Rehabilitation*. 2012; 41:2288–92.
12. Fraser SA, Feldman LS, Stanbridge D, Fried GM. Characterizing the learning curve for a basic laparoscopic drill. *Surgical endoscopy*. 2005; 19(12):1572–8. [PubMed: 16235127]
13. Magill, RA. *Motor Learning and Control: Concepts and Applications*. McGraw-Hill; 2007.
14. Cook JA, Ramsay CR, Fayers P. Using the literature to quantify the learning curve: a case study. *International journal of technology assessment in health care*. 2007; 23(2):255–60. [PubMed: 17493312]
15. Rosser JC, Rosser LE, Savalgi RS. Skill acquisition and assessment for laparoscopic surgery. *Archives of surgery (Chicago, Ill: 1960)*. 1997; 132(2):200–4.
16. Shapiro SJ, Paz-Partlow M, Daykhovsky L, Gordon LA. The use of a modular skills center for the maintenance of laparoscopic skills. *Surgical endoscopy*. 1996; 10(8):816–9. [PubMed: 8694945]
17. Melvin WS, Johnson JA, Ellison EC. Laparoscopic skills enhancement. *American journal of surgery*. 1996; 172(4):377–9. [PubMed: 8873534]
18. McCarter FD, Luchette FA, Molloy M, Hurst JM, Davis K, Johannigman JA, et al. Institutional and Individual Learning Curves for Focused Abdominal Ultrasound for Trauma. *Annals of Surgery*. 2000; 231(5):689–700. [PubMed: 10767790]
19. Schlup MMT, Williams SM, Barbezat GO. ERCP: a review of technical competency and workload in a small unit. *Gastrointestinal Endoscopy*. 1997; 46(1):48–52. [PubMed: 9260705]
20. Van Rij AM, McDonald JR, Pettigrew RA, Putterill MJ, Reddy CK, Wright JJ. Cusum as an aid to early assessment of the surgical trainee. *The British journal of surgery*. 1995; 82(11):1500–3. [PubMed: 8535803]

21. Li JCM, Lo AWI, Hon SSF, Ng SSM, Lee JFY, Leung KL. Institution learning curve of laparoscopic colectomy--a multi-dimensional analysis. *International journal of colorectal disease*. 2012; 27(4):527–33. [PubMed: 22124675]
22. Pendlimari R, Holubar SD, Dozois EJ, Larson DW, Pemberton JH, Cima RR. Technical proficiency in hand-assisted laparoscopic colon and rectal surgery: determining how many cases are required to achieve mastery. *Archives of surgery (Chicago, Ill: 1960)*. 2012; 147(4):317–22.
23. Bartlett A, Parry B. Cusum analysis of trends in operative selection and conversion rates for laparoscopic cholecystectomy. *ANZ journal of surgery*. 2001; 71(8):453–6. [PubMed: 11504287]
24. Fraser SA, Bergman S, Garzon J. Laparoscopic splenectomy: learning curve comparison between benign and malignant disease. *Surgical innovation*. 2012; 19(1):27–32. [PubMed: 21719436]
25. Jaffer U, Cameron AEP. Laparoscopic appendectomy: a junior trainee's learning curve. *JSLs: Journal of the Society of Laparoendoscopic Surgeons/Society of Laparoendoscopic Surgeons*. 2008; 12(3):288–91. [PubMed: 18765054]
26. Kye B-H, Kim J-G, Cho H-M, Kim H-J, Suh Y-J, Chun C-S. Learning curves in laparoscopic right-sided colon cancer surgery: a comparison of first-generation colorectal surgeon to advance laparoscopically trained surgeon. *Journal of laparoendoscopic & advanced surgical techniques Part A*. 2011; 21(9):789–96. [PubMed: 21854205]
27. Okrainec A, Ferri LE, Feldman LS, Fried GM. Defining the learning curve in laparoscopic paraesophageal hernia repair: a CUSUM analysis. *Surgical endoscopy*. 2011; 25(4):1083–7. [PubMed: 20835725]
28. Son G-M, Kim J-G, Lee J-C, Suh Y-J, Cho H-M, Lee Y-S, et al. Multidimensional analysis of the learning curve for laparoscopic rectal cancer surgery. *Journal of laparoendoscopic & advanced surgical techniques Part A*. 2010; 20(7):609–17. [PubMed: 20701545]
29. Murzi M, Cerillo AG, Bevilacqua S, Gilmanov D, Farneti P, Glauber M. 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. *European journal of cardio-thoracic surgery: official journal of the European Association for Cardio-thoracic Surgery*. 2012; 41(6):1242–6. [PubMed: 22232493]
30. Bokhari MB, Patel CB, Ramos-Valadez DI, Ragupathi M, Haas EM. Learning curve for robotic-assisted laparoscopic colorectal surgery. *Surgical endoscopy*. 2011; 25(3):855–60. [PubMed: 20734081]
31. Buchs NC, Pugin Fo, Bucher P, Hagen ME, Chassot G, Koutny-Fong P, et al. Learning curve for robot-assisted Roux-en-Y gastric bypass. *Surgical endoscopy*. 2012; 26(4):1116–21. [PubMed: 22044973]
32. Novotny T, Dvorak M, Staffa R. The learning curve of robot-assisted laparoscopic aortofemoral bypass grafting for aortoiliac occlusive disease. *Journal of vascular surgery: official publication, the Society for Vascular Surgery [and] International Society for Cardiovascular Surgery, North American Chapter*. 2011; 53(2):414–20.
33. Calsina L, Clara A, Vidal-Barraquer F. The use of the CUSUM chart method for surveillance of learning effects and quality of care in endovascular procedures. *European journal of vascular and endovascular surgery: the official journal of the European Society for Vascular Surgery*. 2011; 41(5):679–84. [PubMed: 21333562]
34. Naik VN, Devito I, Halpern SH. Cusum analysis is a useful tool to assess resident proficiency at insertion of labour epidurals. *Canadian journal of anaesthesia*. 2003; 50(7):694–8. [PubMed: 12944444]
35. Lerch L, Donald JC, Olivotto IA, Lesperance M, van der Westhuizen N, Rusnak C, et al. Measuring surgeon performance of sentinel lymph node biopsy in breast cancer treatment by cumulative sum analysis. *The American Journal of Surgery*. 2007; 193(5):556–60.10.1016/j.amjsurg.2007.01.012
36. Kinsey SE, Giles FJ, Holton J. Cusum plotting of temperature charts for assessing antimicrobial treatment in neutropenic patients. *BMJ*. 1989; 299(6702):775–6. [PubMed: 2508918]
37. Papanna R, Biau DJ, Mann LK, Johnson A, Moise KJ Jr. Use of the Learning Curvea “Cumulative Summation test for quantitative and individualized assessment of competency of a surgical

- procedure in obstetrics and gynecology: fetoscopic laser ablation as a model. *American Journal of Obstetrics and Gynecology*. 204(3):218.e1–e9.10.1016/j.ajog.2010.10.910 [PubMed: 21111398]
38. Lane, NE. *Skill Acquisition Rates and Patterns: Issues and Training Implications*. Springer; 1987.
39. Scott DJ, Young WN, Tesfay ST, Frawley WH, Rege RV, Jones DB. Laparoscopic skills training. *American journal of surgery*. 2001; 182(2):137–42. [PubMed: 11574084]
40. Derossis AM, Bothwell J, Sigman HH, Fried GM. The effect of practice on performance in a laparoscopic simulator. *Surgical endoscopy*. 1998; 12(9):1117–20. [PubMed: 9716763]
41. Kolozsvari NO, Kaneva P, Brace C, Chartrand G, Vaillancourt M, Cao J, et al. Mastery versus the standard proficiency target for basic laparoscopic skill training: effect on skill transfer and retention. *Surgical endoscopy*. 2011; 25(7):2063–70. [PubMed: 21614665]
42. Feldman LS, Cao J, Andalib A, Fraser S, Fried GM. A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task: defining “learning plateau” and “learning rate”. *Surgery*. 2009; 146(2):381–6. [PubMed: 19628099]

Figure 1.
Simulators used in this study (A) FLS (B) VBLaST-PT[®]

VBLaST CUSUM Learning Curves - Junior Criterion
(Score = 59, $p_0 = 5\%$)

Figure 2. CUMSUM learning curves for medical students trained on VBLaST-PT[®] using junior criterion success score of 59, acceptable failure rate $p_0 = 5\%$.

FLS CUSUM Learning Curves - Junior Criterion
(Score = 41, $p_0 = 5\%$)

Figure 3. CUMSUM learning curves for medical students trained on FLS using junior criterion success score of 41, acceptable failure rate $p_0 = 5\%$.

VBLaST CUSUM Learning Curves - Intermediate Criterion
(Score = 64, $p_0 = 5\%$)

Figure 4. CUMSUM learning curves for medical students trained on VBLaST-PT[®] using intermediate criterion success score of 64, acceptable failure rate $p_0 = 5\%$.

FLS CUSUM Learning Curves - Intermediate Criterion
(Score = 65, $p_0 = 5\%$)

Figure 5. CUMSUM learning curves for medical students trained on FLS using intermediate criterion success score of 65, acceptable failure rate $p_0 = 5\%$.

VBLaST CUSUM Learning Curves - Senior Criterion
(Score = 68, $p_0 = 5\%$)

Figure 6. CUMSUM learning curves for medical students trained on VBLaST-PT[®] using senior criterion success score of 68, acceptable failure rate $p_0 = 5\%$.

FLS CUSUM Learning Curves - Senior Criterion
(Score = 76, $p_0 = 5\%$)

Figure 7.
CUMSUM learning curves for medical students trained on FLS using senior criterion success score of 76, acceptable failure rate $p_0 = 5\%$.

Figure 8.

CUMSUM learning curves for FLS total and VBLaST-PT[®] total using junior criterion FLS success score of 41, acceptable failure rate $p_0 = 5\%$; VBLaST-PT[®] success score of 59, acceptable failure rate $p_0 = 5\%$.

Figure 9. CUMSUM learning curves for FLS total and VBLaST-PT[®] total using intermediate criterion FLS success score of 65, acceptable failure rate $p_0 = 5\%$; VBLaST-PT[®] success score of 64, acceptable failure rate $p_0 = 5\%$.

Figure 10.

CUMSUM learning curves for FLS total and VBLaST-PT[®] total using senior criterion FLS success score of 76, acceptable failure rate $p_0 = 5\%$; VBLaST-PT[®] success score of 68, acceptable failure rate $p_0 = 5\%$.

Figure 11.
Performance of all the three groups on VBLaST-PT[®] during pre, post and retention tests.

Figure 12.
Performance of all the three groups on FLS during pre, post and retention tests.

Table 1

CUSUM variables for junior, intermediate, and senior success scores

Variable	Value
FLS success score	
Junior	41
Intermediate	65
Senior	76
VBLaST-PT [®] success score	
Junior	59
Intermediate	64
Senior	68
p_0	0.05
$p_1 = 2 \times p_0$	0.10
	0.05
	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
$a = \ln[(1 - s)/s]$	2.77
$b = \ln[(1 - s)/s]$	1.56
$h_0 = -b/(P + Q)$	-2.09
$h_1 = a/(P + Q)$	3.71

Table 2

Sample CUSUM chart calculation for representative subject (medical student no. 8) using VBLaST-PT[®] intermediate criterion score of 64

Trial	Score	Binary Score	CUMSUM Score
1	50.63436	0	0.93
2	4.703992	0	1.86
3	54.44499	0	2.79
4	66.19672	1	2.72
5	79.63605	1	2.65
6	71.40669	1	2.58
7	65.68022	1	2.51
8	75.19932	1	2.44
9	79.51894	1	2.37
10	87.17175	1	2.3
11	79.25769	1	2.23
12	78.9574	1	2.16
13	87.93298	1	2.09
14	75.67827	1	2.02
15	78.34032	1	1.95
16	84.28599	1	1.88
17	91.79767	1	1.81
18	88.03808	1	1.74
19	90.27221	1	1.67
20	93.28559	1	1.6
21	88.21224	1	1.53
22	95.61731	1	1.46
23	96.17134	1	1.39
24	95.25697	1	1.32
25	94.47322	1	1.25
26	96.87401	1	1.18
27	98.93698	1	1.11
28	96.2254	1	1.04
29	97.07671	1	0.97
30	94.5543	1	0.9
31	94.81104	1	0.83
32	102.3512	1	0.76
33	99.7928	1	0.69
34	97.62623	1	0.62
35	99.28382	1	0.55
36	95.04677	1	0.48
37	87.08767	1	0.41
38	102.099	1	0.34

Trial	Score	Binary Score	CUMSUM Score
39	95.46417	1	0.27
40	99.69371	1	0.2
41	101.5044	1	0.13
42	103.162	1	0.06
43	96.90254	1	-0.01
44	103.0554	1	-0.08
45	102.0585	1	-0.15
46	103.2746	1	-0.22
47	100.0405	1	-0.29
48	100.5991	1	-0.36
49	105.9817	1	-0.43
50	106.0223	1	-0.5
51	103.6124	1	-0.57
52	95.19541	1	-0.64
53	106.0763	1	-0.71
54	96.83648	1	-0.78
55	104.0674	1	-0.85
56	105.6214	1	-0.92
57	105.9232	1	-0.99
58	97.11725	1	-1.06
59	103.8196	1	-1.13
60	98.71778	1	-1.2
61	95.21193	1	-1.27
62	108.8464	1	-1.34
63	94.4477	1	-1.41
64	101.9549	1	-1.48
65	105.1214	1	-1.55
66	98.81987	1	-1.62
67	106.8826	1	-1.69
68	98.40398	1	-1.76
69	108.896	1	-1.83
70	97.78539	1	-1.9
71	93.08289	1	-1.97
72	97.81692	1	-2.04
73	99.33787	1	-2.11
74	102.5044	1	-2.18
75	103.4638	1	-2.25
76	100.536	1	-2.32
77	89.39537	1	-2.39
78	102.1981	1	-2.46
79	105.8781	1	-2.53
80	102.1486	1	-2.6

Trial	Score	Binary Score	CUMSUM Score
81	103.1215	1	-2.67
82	103.153	1	-2.74
83	98.1187	1	-2.81
84	83.20196	1	-2.88
85	97.25838	1	-2.95
86	83.99922	1	-3.02
87	98.41599	1	-3.09
88	99.36639	1	-3.16
89	96.70886	1	-3.23
90	90.75867	1	-3.3
91	100.9955	1	-3.37
92	100.3258	1	-3.44
93	106.1349	1	-3.51
94	97.75836	1	-3.58
95	95.34856	1	-3.65
96	107.0357	1	-3.72
97	110.5536	1	-3.79
98	100.2718	1	-3.86
99	106.6799	1	-3.93
100	107.1799	1	-4
101	96.41458	1	-4.07
102	103.7701	1	-4.14
103	106.8826	1	-4.21
104	100.1847	1	-4.28
105	106.4276	1	-4.35
106	105.6214	1	-4.42
107	110.3509	1	-4.49
108	103.3032	1	-4.56
109	108.4005	1	-4.63
110	107.1303	1	-4.7
111	92.97479	1	-4.77
112	90.04699	1	-4.84
113	103.5089	1	-4.91
114	106.6258	1	-4.98
115	93.54834	1	-5.05
116	103.5629	1	-5.12
117	110.3013	1	-5.19
118	108.7383	1	-5.26
119	105.7205	1	-5.33
120	105.6754	1	-5.4
121	107.0312	1	-5.47
122	107.1799	1	-5.54

Trial	Score	Binary Score	CUMSUM Score
123	109.2428	1	-5.61
124	105.3691	1	-5.68
125	110.4995	1	-5.75
126	97.99408	1	-5.82
127	110.5536	1	-5.89
128	112.1616	1	-5.96
129	107.3285	1	-6.03
130	109.1392	1	-6.1
131	102.1531	1	-6.17
132	105.225	1	-6.24
133	108.3915	1	-6.31
134	108.9906	1	-6.38
135	108.8915	1	-6.45
136	110.2518	1	-6.52
137	108.1888	1	-6.59
138	109.9995	1	-6.66
139	102.3392	1	-6.73
140	109.8509	1	-6.8
141	107.3826	1	-6.87
142	110.4995	1	-6.94
143	109.6482	1	-7.01
144	107.5357	1	-7.08
145	110.5491	1	-7.15
146	113.2697	1	-7.22
147	112.4184	1	-7.29
148	112.4139	1	-7.36
149	101.2432	1	-7.43
150	109.9005	1	-7.5

Table 3

Summary of CUSUM learning curve results

Criterion	No. of subjects achieved 5% failure rate	Trial range when 5% failure rate achieved	Overall trial when 5% failure achieved
FLS junior (score = 41)	6	30–59	30
VBLaST-PT [®] junior (score = 59)	5	59–145	87
FLS intermediate (score = 65)	6	30–116	59
VBLaST-PT [®] intermediate (score = 64)	3	73–130	102
FLS senior (score = 76)	3	59–145	102
VBLaST-PT [®] senior (score = 68)	2	73–102	145