Laparoscopic Virtual Reality Training: Are 30 Repetitions Enough?¹

William C. Brunner, M.D., James R. Korndorffer, Jr., M.D., Rafael Sierra, M.D., Nader N. Massarweh, B.S., J. Bruce Dunne, Ph.D., C. Lillian Yau, Ph.D., and Daniel J. Scott, M.D.²

Tulane Center for Minimally Invasive Surgery, Tulane University Health Sciences Center, New Orleans, Louisiana

Submitted for publication December 12, 2003

Background. Current literature suggests that novices reach a plateau after two to seven trials when training on the MIST VR laparoscopic virtual reality system. We hypothesize that significant benefit may be gained through additional training.

Materials and methods. Second-year medical students (n = 12) voluntarily enrolled under an IRB-approved protocol for MIST VR training. All subjects completed pre- and posttraining questionnaires and performed 30 repetitions of 12 tasks. Performance data were automatically recorded for each trial. Learning curves for each task were generated by fitting spline curves to the mean overall scores for each repetition. Scores were assessed for plateaus by repeated measures, slope, and best score.

Results. On average, subjects completed training in 7.1 h. (range, 5.9–9.2). Two to seven performance plateaus were identified for each of the 12 MIST VR tasks. Initial plateaus were found for all tasks by the 8th repetition; however, ultimate plateaus were not reached until 21–29 repetitions. Overall best score was reached between 20 and 30 repetitions and occurred beyond the ultimate plateau for 9 tasks.

Conclusions. These data indicate that a lengthy learning curve exists for novices and may be seen throughout 30 repetitions and possibly beyond. Performance plateaus may not reliably determine training endpoints. We conclude that a significant and variable amount of training may be required to achieve maximal benefit. Neither a predetermined training duration nor an arbitrary number of repetitions may be adequate to ensure laparoscopic proficiency following simulator training. Standards which define performance-based endpoints should be established.

Key Words: laparoscopic surgery; virtual reality; laparoscopic skills training; surgical education; learning curve; MIST VR; competency.

INTRODUCTION

Rising costs associated with traditional surgical training in the clinical setting as well as an emphasis on reduction of patient risk have promoted the development of virtual reality training adjuncts for surgical education, particularly involving minimally invasive surgery [1–3]. Laparoscopic simulators have proven effective in providing skills which are transferable to the operating room environment [4–6]. The unique features of laparoscopic surgery, i.e., fixed access points, the fulcrum effect [7, 8], two-dimensional representation of a three-dimensional environment [9], and long instruments with diminished tactile feedback create a significant barrier to technical mastery.

The Minimally Invasive Surgical Trainer – Virtual Reality or MIST VR (Mentice, Inc., San Diego, CA) is the most widely used computer-based system, as 40 of 253 General Surgery residency programs in the United States have at least one unit. Although the MIST VR has been validated by several studies, no consensus exists on training duration or curriculum design [10–20].

Definitions for plateau as a means of assessing training endpoints have been vague, widely variable, and cumbersome for the task of real-time training assessment. Nonetheless, some studies suggest that a plateau in training may be reached after as few as two to seven repetitions of an individual task [21, 22]. If the learning curve is in fact this short, significant monetary investment in these training systems may not be warranted since costs ($20,000 list price) would greatly outweigh the benefit of such brief training. Furthermore, data from other training systems and clinical practice suggest that a more prolonged learning curve...
may exist [23, 24]. The purpose of this study was to determine the learning curves for MIST VR training in an effort to quantify the amount of training that may be appropriate.

**MATERIALS AND METHODS**

Second-year medical students \( n = 12 \) were recruited from the Rudolph Matas Surgical Society, a medical student surgical interest group at Tulane University School of Medicine, and were voluntarily enrolled under an IRB-approved protocol for MIST VR training. Each subject completed a pretraining questionnaire to collect demographic data and assess for prior exposure and baseline laparoscopic experience. All subjects were trained on one of three identical MIST VR stations in the Tulane Center for Minimally Invasive Surgery Simulation and Training Laboratory (Fig. 1). Training was self-paced and accomplished on successive days as students' schedules permitted.

The MIST VR system creates a virtual laparoscopic trainer using a Pentium© class personal computer, a video monitor, and a laparoscopic interface (Immersion Medical, Gaithersburg, MD) containing two pistol-grip instrument handles and an electrocautery foot pedal. Twelve basic laparoscopic tasks from Core Skills 1 and Core Skills 2 may be practiced, as previously described [25]. The computer automatically collects data concerning time, errors, economy of motion, and economy of diathermy and generates an overall score for each trial.

All subjects were given a brief introduction to the system, including task design and avoidance of errors. The computer-generated
tutorials were demonstrated for each task. The level of difficulty was set to the “Easy” default settings for all training. A student coordinator and two research assistants were available for assistance and questions, although no active instruction was given. Each subject initially completed 20 consecutive repetitions of each of the 12 tasks. Interim analysis suggested a lack of plateau on some tasks and all subjects were asked to complete an additional 10 repetitions of all 12 tasks. All subjects completed a posttraining questionnaire regarding their impressions about training.

Learning curves for each task were generated by fitting spline curves to the mean overall score for each repetition. A second plot using a five-trial moving average was generated to reduce trial-to-trial variability of individual data points and better demonstrate performance trends [26]. Data analysis was carried out using three methods: (1) repeated measures, (2) learning curve slope, and (3) best overall score. Ultimate plateau was defined as the best plateau achieved within 30 repetitions.

Repeated measures using mixed models were compared for trials 1–10, 11–20, and 21–30 using raw data [27]. Plateaus were defined as no statistical difference between clusters of 10 trials using S-Plus software (MathSoft, Inc., Seattle, WA).

Since the repeated measures analysis showed only the cluster of 10 trials where a plateau was thought to have occurred, further data evaluation was undertaken to identify a more finite plateau location. Using the moving average curve, a 2-point slope was determined between consecutive repetitions. An ideal slope for a performance plateau would be zero; therefore, we defined a reasonable estimate for plateau as a slope of greater than −0.1. Moving averages were calculated using Microsoft Excel (Microsoft, Inc., Redmond, WA) and slope analysis was carried out using Sigma Stat and Sigma Plot (SPSS, Inc., Chicago, IL).

In an effort to locate the point of maximal training benefit received within 30 repetitions the best overall score for each task was defined as the repetition with the lowest mean score.

**RESULTS**

There were four women and eight men, ranging in age from 24 to 31 years. All 12 subjects completed the pretraining questionnaire, 30 repetitions of each of the 12 assigned tasks, and the posttraining questionnaire. Time for completion of all 12 tasks ranged from 5.9 to 9.2 h (mean, 7.1). Results of the data analyses are detailed in Table 1. Graphs of the overall score and moving average curves are shown for Core Skills 1 (Figs. 2–4) and Core Skills 2 (Fig. 5–7).

Repeated measures using mixed models showed that a statistical plateau was reached in the first 10 repetitions for 1 task only (Withdraw Insert). A plateau was achieved within the second 10 repetitions for 10 tasks (Acquire Place, Transfer Place, Traverse, Diathermy, SC Stretch, SC Clip, SC Stretch Clip, SD Stretch, SD Diathermy, and SD Stretch Diathermy). However, in 6 of these 10 tasks (Acquire Place, Transfer Place, SC Stretch Clip, SD Stretch, SD Diathermy, and SD Stretch Diathermy), a second plateau was achieved in the final 10 trials. A plateau was first reached in the final 10 trials for 1 task only (Manipulate Diathermy).

**TABLE 1**

<table>
<thead>
<tr>
<th>Task</th>
<th>Plateau analysis method</th>
<th>Repeated measures</th>
<th>Slope</th>
<th>Best score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquire Place</td>
<td>Repeated measures</td>
<td>21–30</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Transfer Place</td>
<td>Repeated measures</td>
<td>21–30</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Traverse</td>
<td>Repeated measures</td>
<td>11–20</td>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td>Withdraw Insert</td>
<td>Repeated measures</td>
<td>1–10</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>Diathermy</td>
<td>Repeated measures</td>
<td>11–20</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>Manipulate Diathermy</td>
<td>Repeated measures</td>
<td>21–30</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td>SC Stretch</td>
<td>Repeated measures</td>
<td>11–20</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>SC Clip</td>
<td>Repeated measures</td>
<td>11–20</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>SC Stretch Clip</td>
<td>Repeated measures</td>
<td>21–30</td>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td>SD Stretch</td>
<td>Repeated measures</td>
<td>21–30</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>SD Diathermy</td>
<td>Repeated measures</td>
<td>21–30</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>SD Stretch Diathermy</td>
<td>Repeated measures</td>
<td>21–30</td>
<td>29</td>
<td>29</td>
</tr>
</tbody>
</table>
Slope analysis demonstrated an initial plateau by the 8th repetition in all 12 tasks. Multiple (2–7) plateaus were identified for each task with the ultimate plateaus occurring between repetitions 21 and 29.

Best overall score occurred at 20 for 1 task, between repetitions 27 and 29 for 4 tasks, and at repetition 30 for 7 tasks.

All participants noted subjective improvement in their laparoscopic technical skills following training. Laparoscopic self-rating scores improved by 1.6 points according to a 5-point Lickert scale ($P < 0.05$). All participants described an improvement in their comfort level and reported that the training program met their expectations. Seven of 12 subjects expressed interest in additional advanced laparoscopic training curricula in suturing and ultrasound.

**DISCUSSION**

As new simulators for skills enhancement are developed and implemented, one important question is “how much training is enough?” Previous investigations have suggested that performance plateaus derived from learning curves may help educators determine appropriate training durations [23]. For the MIST VR system, plateaus have been identified at two to seven repetitions, but methods for defining plateaus have relied upon comparisons of individual trials during very brief periods of training [21, 22]. These data may

**FIG. 3.** Performance data for Core Skills 1 Traversal and Withdraw Insert.

**FIG. 4.** Performance data for Core Skills 1 Diathermy and Manipulate Diathermy.
be flawed due to issues concerning study design, sample size, and statistical methodology. Because of the lack of literature describing valid ways of determining plateaus, we explored three methods and applied them to a relatively large, homogeneous sample, during a lengthy training protocol.

Repeated measures are a robust method for determining differences in performance between groups [27]. When they were applied to our data, we detected an initial plateau during repetitions 1–10, 11–20, and 21–30 for 1, 10, and 1 tasks, respectively. However, ultimate plateaus were detected during repetitions 21–30 for 8 of 12 tasks. Although valid, the repeated measures analysis was limited by only identifying general areas where a plateau occurred. By plotting 5-point moving averages, meaningful learning curves were generated and a more sensitive slope analysis was performed. According to slope, an initial plateau was identified for all tasks by eight repetitions, consistent with results from other studies [21, 22]. However, with additional training there was evidence of continued benefit, since numerous plateaus were identified and an ultimate plateau did not occur until repetitions 21–29. According to best score, maximal performance was achieved during repetitions 27–30 for 11 tasks. Moreover, a best score was detected beyond the ultimate slope plateau for 9 tasks, suggesting that additional plateaus might have occurred had training continued longer.

This study has several important findings. First, identifying plateaus using statistically valid methods is cumbersome and not practical for providing endpoints for training on a real-time basis. Second, plateaus may not reliably predict appropriate training

![FIG. 5. Performance data for Core Skills 2 SC Stretch and SC Clip.](image)

![FIG. 6. Performance data for Core Skills 2 SC Stretch Clip and SD Stretch.](image)
endpoints at all. As seen graphically from the standard deviations in scores (Figs. 2–7), there was significant variability in individual ability, even among laparoscopically naïve medical students. Curricula based upon an arbitrary duration or number of repetitions may be inefficient since some participants will be overtrained and kept longer than necessary. Similarly, such curricula may not be uniformly effective, since some participants will be undertrained unless additional practice is afforded.

Since the ultimate goal of simulator-based skills training is to foster the development of proficiency, a new paradigm seems in order. Instead of a “one size fits all” approach, curricula should be based on achieving an appropriate level of performance. By establishing performance standards, training can be efficiently tailored to each individual’s needs. For those with superb innate ability, minimal practice may be required and time can be spent on other activities. For those who require more practice, appropriately lengthy sessions may be scheduled until a predetermined proficiency level is reached. Performance-based curricula may significantly enhance the efficiency and efficacy of skills training and are currently being explored.

In conclusion, this study clearly demonstrates that a lengthy learning curve exists for all MIST VR tasks for novices. Plateaus indicate that significant benefit occurs during 30 repetitions and possibly beyond, but they do not reliably establish endpoints for training. To further improve curricula and to ensure proficiency, goal-directed training to performance-based standards should be implemented.

ACKNOWLEDGMENT

Funding for this study was provided in part through an unrestricted educational grant from Ethicon Endo-Surgery, Inc., Cincinnati, Ohio.

REFERENCES


