Comparison of Training on Two Laparoscopic Simulators and Assessment of Skills Transfer to Surgical Performance

Patricia L Youngblood, MEd, PhD, Sakti Srivastava, MBBS, MS, Myriam Curet, MD, FACS, Wm LeRoy Heinrichs, MD, PhD, Parvati Dev, PhD, Sherry M Wren, MD, FACS

BACKGROUND: Several studies have investigated the transfer of surgical trainees' skills acquired on surgical simulators to the operating room setting. The purpose of this study was to compare the effectiveness of two laparoscopic surgery simulators by assessing the transfer of skills learned on simulators to closely matched surgical tasks in the animal laboratory.

STUDY DESIGN: In this post-test-only Control group study design, 46 surgically naive medical student volunteers were randomly assigned to one of three groups: Tower Trainer group (n = 16), LapSim group (n = 17), and Control group (n = 13). Outcomes measures included both time and accuracy scores on three laparoscopic tasks (Task 1: Grasp and Place; Task 2: Run the Bowel; Task 3: Clip and Cut) performed on live anesthetized pigs, and a global rating of overall performance as judged by four experienced surgeons.

RESULTS: The Tower Trainer group performed significantly better than the Control group on 1 of 7 outcomes measures—Task 3: Time (p < 0.032), although the LapSim group performed significantly better than the Control group on 2 of 7 measures—Task 3: Time (p < 0.008) and Global score (p < 0.005). In comparing the two simulators, the LapSim group performed significantly better than the Tower Trainer group on 3 of 7 outcomes measures—Task 2: Time (p < 0.032), Task 2: Accuracy (p < 0.030) and Global score (p < 0.005), although the Tower Trainer group did not perform significantly better than the LapSim group on any measure.

CONCLUSIONS: This study demonstrated that naïve subjects trained on a virtual-reality part-task trainer performed better on live surgical tasks in a porcine model as compared with those trained with a traditional box trainer. These findings could aid in selection of appropriate training methodologies. (J Am Coll Surg 2005;200:546–551. © 2005 by the American College of Surgeons)

Recent studies of simulation-based training for laparoscopic surgery have focused on construct validity1–4 and transfer of skills to the operating room environment.5–8 Currently available simulators range from skills-specific part-task trainers9 to electronically enabled mannequins in an operating room setting.10 The part-task trainers include both physical simulators, such as box trainers,11 and computer-aided, virtual-reality (VR) systems, incorporating complex computer models of instrument-tissue interactions.12

This study compares the efficacy of two part-task skills trainers for preparing trainees to perform laparoscopic procedures. The first is a VR-based system (LapSim Basic Training program; Surgical-Science AB) and the second is a physical simulator (Tower Trainer; Simu-labs). Both systems have advantages and disadvantages.13 LapSim has more realistic graphics and the orientation of workspace to monitor is similar to the actual operating room environment, but Tower Trainer has been used for many years in surgical training programs and incorporates use of real laparoscopic instruments. In this study, our first aim was to assess whether skills acquired with either simulator are transferable to live surgical pro-
cures by comparing the performance in the animal laboratory of those trained on either simulator with those who received no training. Our second aim was to assess whether training with LapSim was less effective, equally effective, or more effective than training with the Tower Trainer.

METHODS
Research subjects
In this post-test-only Control group design, volunteer subjects were recruited from the pool of first and second year (preclinical) medical students, who had no earlier experience performing laparoscopic surgery, as assessed by recruitment screening questionnaire. Subjects were sequentially randomized to one of three groups: LapSim training group, Tower Trainer training group, and Control group (no training). A total of 46 subjects participated in the study: 17 in the LapSim training group, 16 in the Tower Trainer training group, and 13 in the Control group. There were 32 men and 14 women in the study. Of the 33 who received training, only 1 was left-handed. All subjects completed an IRB consent form and on completion were given a gift certificate to the University bookstore as incentive for their participation. Each subject completed training sessions and assessments within a 12-day period.

Description of the two simulators and training tasks
In the LapSim system, users manipulate modified surgical instrument handles that activate virtual instrument tips seen on the computer screen. The input device is a bimanual force feedback surgical workstation codveloped by Stanford University Medical Media and Information Technologies (SUMMIT) and Immersion Medical. Three of LapSim’s seven basic skill training exercises were used in this study—grasping and placing; grasping and cutting; and clip-applying and cutting, none of which incorporates haptic feedback.

In the Tower Trainer system, users look into a periscope to manipulate objects on an angled stage inside the simulator. Users manipulate real surgical instruments—grasper, clip-applier, and scissors—inserted into ports at each side of the simulator. The three training tasks for this simulator were: grasping and placing; grasping and cutting a premarked string; and grasping, clip-applying, and cutting a premarked string. In the first task, dried beans simulated gallstones, in the second task string simulated the bowel, and in the third task, string simulated a vessel.

Study protocol
Subjects completed 4 researcher supervised, 45-minute training sessions. A 30-minute final assessment session in the porcine laboratory was completed within a 2-week period. Subjects had up to 45 minutes to complete the training protocol for both simulators. Subjects were initially oriented to the equipment and specific tasks; no additional instruction was provided. A nonsurgeon researcher supervised each training session. At each training session subjects completed 10 repetitions of all 3 training tasks corresponding to the surgical tasks of grasping and placing gallstones in an endo bag, running the bowel, and clipping and cutting an artery. Subjects in the Control group received no training but were scheduled randomly among other subjects for completing the three surgical tasks in the animal surgery. The training laboratory was not accessible to subjects outside of the subjects’ scheduled training periods.

Final assessment in the operating room
For their final assessment session, subjects were gowned and gloved and given brief instructions by a surgeon researcher in the use of the surgical instruments. No prompting or guidance was given to any subject during the operative assessment. Before each session, two experienced laparoscopic surgeons prepared the surgical field by placing the trocars. For Task 1, they placed 18 to 20 white beans on the liver surface and inserted an Endocatch (Auto Suture) basket nearby for the grasping and placing task. For Task 2, the surgeons placed six dark-colored “marker stitches” through the bowel at approximately 8-cm intervals to serve as target sites for the grasping task. For Task 3, the team dissected a mesenteric artery and vein as target structures for the clipping and cutting task. Each subject’s performance on these structured assessment tasks was videotaped endoscopically for blinded scoring at a later time. The blinded performance of two expert laparoscopic surgeons was also recorded and rated by the panel of four experienced surgeons, for a benchmark comparison.

Blinded scoring of videotapes
Each subject’s videotaped performance was independently rated by four experienced surgeons. For each of
the three tasks, the surgeons provided two assessments: time to completion and an accuracy score derived from dividing number of correct manipulations by number of attempts. In addition, surgeons gave each subject a Global score on a scale of 1 (poor) to 5 (excellent). Before scoring, all surgeons agreed to a common, behaviorally anchored rating scale. Videotape scoring protocol is presented in Table 1. Inter-rater reliability was established by calculating Spearman Rho correlation coefficients for each pair of raters. These correlation coefficients were as follows: 0.78, 0.86, 0.83, 0.88, 0.83, and 0.88 (each was significant at p < 0.01).

**Statistical methods**

Performance data from the final assessment in the animal surgery were analyzed using a one-way ANOVA with group membership as the independent variable and subjects’ scores on each of three surgical tasks and Global scores as dependent variables. When significance at the p < 0.05 level was found, Tukey’s post hoc comparison was performed to compare pairs of means within the ANOVA. Global scores were analyzed using the Kruskal-Wallis test. Data were analyzed using SPSS for Windows (SPSS Inc).

**RESULTS**

Table 2 provides a summary of the results for each group on the three surgical tasks and the Global score that were used as outcomes measures in this study. Figures 1 to 3 show comparison of the time scores for each group on each of three surgical tasks, comparison of the accuracy scores for each group for each of the three tasks, and comparison of the Global scores for each treatment group.

**Time scores**

These analyses show that with regard to time scores, the VR-trained group performed significantly better than the Control group on two of the three tasks and significantly better than the Tower Trainer group on Task 2: Running the Bowel. On the other hand, the Tower Trainer–trained group performed significantly better than the Control group on only one task—Task 3: Clipping and Cutting. The Tower Trainer group’s time scores were not significantly better than those of the VR-based...
Figures

Figure 1. Comparison of mean accuracy scores by group. **Significance over other treatment group. **Run the Bowel: LapSim mean score versus Tower Trainer mean score, p < 0.030.

Figure 2. Comparison of mean Global ratings (1–5) for each group. *LapSim mean score versus Control mean score, p < 0.005. **LapSim mean score versus Tower Trainer mean score, p = 0.005.

**DISCUSSION**

Previously reported transfer validity studies indicate that earlier “part-task” training is associated with improved performance of medical students in animal laboratories and surgical residents in operating rooms. Few studies have incorporated a true randomized controlled trial in which the operating room performances of subjects trained on two types of laparoscopic surgical skill simulators were compared with the performance of a Control group that received no training.

Face validity of each simulator indicates inherent strengths for both in preparing trainees for laparoscopic surgery. For example, in the Tower Trainer simulator, trainees manipulate real objects with real surgical instruments, giving the user an authentic haptic experience. In the LapSim simulator, trainees view realistic graphics of the anatomic region while manipulating virtual tissue with virtual instruments. We hypothesized that those trained by either method would perform better on the surgical tasks in the animal laboratory than those who received no training (Control group). We were surprised to find that the performance of the Tower Trainer group was significantly better than the Control group on only one of seven outcomes measures (time score for Task 3). This is in contrast to recent findings of Munz and colleagues, on a similar study comparing a classic box trainer with the LapSim VR simulator. One possible explanation for the apparent lack of transfer for those trained with the Tower Trainer in our study could be the simplistic representation of the tissues and training tasks. Findings for those trained with the LapSim supported our hypothesis that training with a simulator would result in better performance of the surgical tasks than those who received no training (Control group). Indeed, those who trained with the LapSim performed significantly better than the Control group on two of the seven outcomes measures: time score for Task 3 and Global score.

An additional goal of the study was to compare the two simulators. The Tower Trainer is a low-cost box trainer that has been used widely for many years in surgical training programs to prepare residents for performing laparoscopic surgery. The LapSim is a more expensive computer-based training technology that appears to provide greater realism in that users report a greater sense of immersion in the simulated surgical tasks. In contrast to the findings of Munz and colleagues, our results show that the LapSim group performed significantly better than the Tower Trainer group on three of the seven...
outcomes measures—Time and Accuracy scores for Task 2 (Run the Bowel) and the Global score. It is noteworthy that the Tower Trainer group did not perform significantly better than the LapSim group on any of the measures. This clearly shows that the LapSim simulator affords a more effective and efficient training experience than does the Tower Trainer simulator. The question remains whether the additional benefit is worth the additional cost of the computer-based system.

Differences in performance of the two groups may be explained by noting the unique characteristics of both systems. For example, in Task 2: Running the Bowel, the Tower Trainer subjects were asked to grasp and manipulate a piece of twine, which was laid on a flat inclined stage and had minimal multidirectional movement. In contrast, the simulated tissue in the LapSim simulator had multidirectional, dynamic motion. These properties in the LapSim simulator may have led to better transfer when subjects were asked to manipulate the small bowel, which also has multidirectional, dynamic motion. In addition, placement of objects on a flat, inclined stage prevented subjects from “past pointing” in the “z” or forward dimension. In contrast, the LapSim subjects were able to “past point” in all dimensions, which more accurately reflects the surgical environment.

Another factor that may explain why LapSim subjects performed better on Task 2: Run the Bowel, is that this task required true two-handed coordination. In the Tower Trainer group there was no coordination between the two hands other than simple stabilization with the first hand to allow the second hand to perform the task. In contrast the LapSim subjects had to control the amount of tension in the grasping hand or the tissue would rupture.

It seems that absence of any significant difference on subjects’ performance on surgical Task 1: Pick and Place, suggests that both simulators provided adequate practice of a simple one-handed task. This may indicate that use of a low-fidelity simulator is adequate for simple tasks and the VR simulator appeared to be more effective for learning more complex, two-handed, coordinated tasks. Our findings suggest that decisions about which systems should be used are highly dependent on the nature of the surgical task—its complexity (how many steps and type of hand coordination), need for haptic feedback, and nature of the visual-spatial field.

There was no significant difference between the two simulator groups on trainees’ performance on Task 3: Clipping and Cutting, although both groups performed significantly better than the Control group on the time score. One possible explanation for this is a ceiling effect. Both groups were able to reach the criterion level of performance after four training sessions.

It is important to note that with the small sample sizes in this study there is an increased likelihood of a type II error. This means there may have been significant differences between groups that were not detected because of the limited power of the study. When the total sample size across the three groups is 46, distributed across the groups as specified, a one-way ANOVA will have 80% power to detect at the 0.05 level an effect size of approximately 1.5.

In conclusion, this study clearly highlights the need for task-specific comparisons of surgical simulators. In this randomized, controlled, double-blinded study, the training tasks were carefully matched to each other and to the surgical tasks performed in the animal laboratory. Future studies of this type are needed to better understand the relative strengths and limitations of each type of simulator. Findings from this and future evaluation research on simulators may provide guidance to residency program directors who seek objective data on efficacy and cost effectiveness of using surgical simulators to enhance laparoscopic skill development in their training programs.15,16

Author Contributions
Study conception and design: Youngblood, Srivastava, Curet, Heinrichs, Dev, Wren
Acquisition of data: Youngblood, Srivastava, Curet, Heinrichs, Wren
Analysis and interpretation of data: Youngblood, Srivastava, Curet, Heinrichs, Wren
Drafting of manuscript: Youngblood, Srivastava, Curet, Heinrichs, Wren
Critical revision: Youngblood, Srivastava, Curet, Heinrichs, Wren
Statistical expertise: Youngblood
Obtaining funding: Curet, Heinrichs, Dev, Wren
Supervision: Heinrichs

Acknowledgment: We would like to thank Yamil Saenz, VET EDU Specialist, for his supervision of the surgical tasks in the Stanford University School of Medicine animal laboratory, and Shannon Moffett and Rina Dutta for their assistance in recruiting students and monitoring training sessions. We would also like to thank Robert Cheng for technical assistance with the surgical simulators.
REFERENCES


COLLABORATORS

CHICAGO

C. S. BACON, M. D., President Chicago Medical Society

D. A. K. STEELE, M. D., President Chicago Surgical Society

FRANK T. ANDREWS, M. D., President Chicago Gynecological Society

NEW YORK

W. H. Alport, M. D.

Carl Beck, M. D.

Arthur D. Bevac, M. D.

James B Dwy, M. D.

Henry T. Byrd, M. D.

Frank Cary, M. D.

Claron Davidson, M. D.

Joseph B. Delee, M. D.

E. J. Doering, M. D.

Frank D. Eto, M. D.

Alex H. Ferguson, M. D.

Albert Goldsmit, M. D.

A. E. Halitazid, M. D.

M. L. Harms, M. D.

Wu, M. Harmony, M. D.

Freda Hendron, M. D.

J. C. Hung, M. D.

Henry F. Lewis, M. D.

L. L. McAdams, M. D.

Andrew McDaniel, M. D.

J. P. Newman, M. D.

A. J. Oehserer, M. D.

John F. Owen, M. D.

C. E. Paddock, M. D.

S. C. Plasmon, M. D.

C. B. Reed, M. D.

H. M. Richardson, M. D.

E. C. Riebel, M. D.

Emil Ries, M. D.

Bryan Robinson, M. D.

Wm. E. Schreuder, M. D.

E. J. Senn, M. D.

Walter Van Hool, M. D.

Berta Van Hooven, M. D.

Lucy Walde, M. D.

T. J. Watkins, M. D.

BOSTON

J. D. Eustis, M. D.

Wm. T. Bell, M. D.

Austin Bill, Jr., M. D.

PHILADELPHIA

Maurice H. Richardson, M. D.

Walter L. Burre, M. D.

Edward M. Stroebel, M. D.

E. E. Montgomery, M. D.

John B. Denler, M. D.

Edward P. Davis, M. D.

BALTIMORE

Howard A. Kelly, M. D.

J. C. Bloodgood, M. D.

J. Whitley Williams, M. D.

SAN FRANCISCO

Beverly MacManus, M. D.

George Goodell, M. D.

DENVER

W. W. Grant, M. D.

Carey Fleming, M. D.

WASHINGTON

Henry D. Fry, M. D.

J. Wesley Brown, M. D.

ROCHESTER, MINN.

William J. Mayo, M. D.

Charles H. Mayo, M. D.

ST. LOUIS

N. B. Carson, M. D.

Walter B. Demott, M. D.

KANSAS CITY

A. H. Cordier, M. D.

J. D. Graham, M. D.

LOUISVILLE

Lewis S. McMurtry, M. D.

A. Morgan Vauter, M. D.