EVALUATION OF PHYSICAL VERSUS VIRTUAL SURGICAL TRAINING SIMULATORS

[1]Waxberg, S.L., [2]Goodell, K.H., [1]Avgerinos, D.V.,
[2]Schwaitzberg, S.D., [1]Cao, C.G.L.

[1] Department of Mechanical Engineering, School of Engineering,
Tufts University, Medford, MA
[2] Department of Surgery, Center of Minimally Invasive Surgery,
Tufts-New England Medical Center, Boston, MA

ABSTRACT

With the recent attention on patient safety, there is an increased interest in standardized training for laparoscopic surgeons. Studies have shown that laparoscopic simulators can be used to train surgical skills. A comparison of two popular systems (a real physical model and a virtual model) was conducted to determine the relative effectiveness of the systems for training purposes. Twenty-two medical students and surgical residents were tested on both simulators. Time to task completion and errors committed were recorded and compared. Our results showed that the physical training system was more sensitive to the experience levels of the subjects than the virtual system, and may be more effective as a tool for standardized training. However, as virtual reality technology becomes better developed, and surgeons become more familiar with the technology, we may see a change.

INTRODUCTION

For many common surgical operations, laparoscopic techniques have proven superior over the open technique. Benefits include decreases in recovery time, scarring, and rate of post-operative infection. Increased demand for laparoscopic surgery requires that *more* surgeons become trained, and that they are *better* trained in the technique. In a departure from the traditional apprenticeship model of using animals and human cadavers for practice, surgical training is moving towards using standardized simulation modules as well as virtual reality systems (Hoffman & Dzung, 1997). Several recent studies have demonstrated that physical and virtual reality laparoscopic simulators are effective in teaching skills that are transferable to real laparoscopic tasks (Munz, Kumar, Moorthy, Bann, & Darzi, 2004; Fried, Derossis, Bothwell, & Sigman, 1999; Gallagher, Richie, McClure, & McGuigan, 2001; Derossis, Fried, Abrahamowicz, Sigman, Barkun, & Meakins, 1998), and that training with surgical simulators can reduce risks and the lack of cognitive and visuomotor skills thereby improving performance in the operating room (Seymour, Gallagher, Roman, O'Brien, Bansal, Andersen, & Satava, 2002).

Surgical skill, as with any other motor skill, is acquired and mastered after going through various stages or phases (Fitts, 1964): 1) the cognitive phase teaches a novice the process to be performed and how best to attempt the first few trials, 2) the associative phase teaches the subject to make more subtle adjustments in how the skill is performed, and 3) the autonomous phase, which begins after months or even years of practice, consists of the time that the experienced person possesses a skill that has become largely automatic (Schmidt, 1998). As a result, the process of learning and mastering of laparoscopic skills demands great time and effort. Simulators, offering availability and a relatively safe environment removed from the stress of operating on a real patient, are the ideal tools for learning and mastering surgical skills.

Currently, there are two popular training simulators which are used to investigate the effectiveness and validity of this training approach. Studies have shown that one of these models, the virtual reality simulator MIST-VR (Minimally Invasive Surgical Trainer-Virtual Reality), is an effective training tool in that it can distinguish between novice, junior, and experienced surgeons in several basic surgical skills (Gallagher et al, 2001). Similarly, the simulator MISTELS (McGill Inanimate System for Training and Evaluation of Laparoscopic Skills), a physical box trainer, has also been shown to be able to discriminate between experience levels (Derossis et al, 1998). In addition, training on the MISTELS can be correlated to improvements, in vivo, on laparoscopic pig procedures (Fried et al, 1999). The Society of American Gastrointestinal Endoscopic Surgeons (SAGES) has also accepted this system as its official trainer for certification purposes (Peters, Fried, Swanstrom, Soper, Sillin, Schirmer, Hoffman, & SAGES FLS Committee, 2004). These two types of trainers were designed to measure motor skill performance using very different motor tasks, all of which contain some combination of the basic surgical skills that are used in laparoscopic surgery. Some tasks are more representative of the actual surgical task, such as the suturing and knot tying tasks, while others are less representative, such as the peg-transfer task. The systems require that users perform different types of tasks, and it is not known how the virtual reality simulators compare to the real/physical box trainers.

The first step towards establishing the usefulness of a trainer is ensuring its ability to consistently and reliably differentiate between ability levels. Once the usefulness of the system has been established, it is then possible to determine what kind of trainer, real or virtual, can best train surgeons at improving their surgical skills. This study was designed to compare the relative effectiveness of the MIST-VR system and the MISTELS physical trainer box at differentiating surgical skill levels. To this end, medical students and residents were tested on both laparoscopic simulators. Each system was used as it was designed, without modifications to the tasks or the testing conditions, in order to evaluate its usefulness as a training system. In other words, we did not attempt to compare the two systems in terms of their tasks.

METHODS

Subjects:

Twenty-two participants (5 medical students, 7 interns, 2 second-year residents, 3 third-year residents, 3 fourth-year residents, and 2 fellows) from Tufts-New England Medical Center, Stamford Hospital, and Tufts Medical School participated in this study. All interns, residents and fellows were from the Departments of Surgery of the two hospitals. Participants volunteered for the study and were highly motivated. The study protocol was approved by the local ethics committee and participants gave their informed consent form.

Apparatus:

The MIST-VR system is comprised of a 2.8 GHz Pentium 4 PC, a 13-inch computer monitor, and a laparoscopic tool base with two tools each having 5 degrees of freedom. The monitor was placed at eye level and the tool base is positioned at waist level directly in front of the participant.

The MISTELS system consists of a box frame, laparoscopic tools, two trocars, an opaque plastic cover, an endoscopic camera with a light source, and a 13-inch TV monitor. The monitor was placed to the right of the participant. The laparoscopic tools and camera were inserted into the box to simulate laparoscopic surgery.

Tasks:

Participants began each session by filling out a short questionnaire with biographic and surgical training information.

Virtual Simulator Tasks. Each MIST-VR session was made up of 13 tasks. Before each task, a demonstration of that task was shown accompanied by a verbal explanation. The 13 tasks were:

1. Acquire Place: a ball was picked up with one tool and placed in a box.
2. Transfer Place: a ball was picked up, transferred to other tool, and placed in a box.
3. Traversal: a cylinder was grabbed in a specified section and traversed down, alternating tools.
4. Withdraw Insert: a ball was picked up with one tool, touched with the second tool, the second tool was removed from the screen, reinserted again and then the ball was touched again while inside a box.
5. Diathermy: point and touch a box fixed to a ball with working tool.
6. Manipulation Diathermy: combined tasks 4 and 5.
7. SD Stretch: highlighted end of a barbell was grasped and stretched until there was a change in color, indicating the correct length.
8. SC Clip: highlighted section of a barbell was clipped.
9. SC Stretch Clip: combined tasks 7 and 8.
10. SD Diathermy: task 5 on the shaft of a barbell moving from fixed to highlighted end.
11. SD Stretch Diathermy: combined tasks 5 and 7.
12. Start Stitch: a needle was grabbed and pushed through tissue at a specified location.
13. Half Square Knot: a needle was grasped with the right tool, a thread was wound around the left tool shaft, and then the ends were tightened to create a knot.

Real Box Trainer Tasks. Participants were first shown an introduction video that summarized the basics of laparoscopic surgery. They were given an overview of all tasks, as well as in-depth description of the task objectives. The five tasks were:

1. Peg Transfer: pegs were acquired from one pegboard using one tool, transferred to the other tool, and then

placed on a new pegboard and then repeated in reverse.

2. Cutting Pattern: Endoscopic scissors cut out a prescribed circle from a piece of gauze.
3. Endoloop: a ligating loop was tightened around a specified section of a sponge appendage.
4. Intra-Corporeal Knot: a stitch was made through marked spots on a Penrose drain and then three double-throw knots are tied and secured.
5. Extra-Corporeal Knot: after the stitch was made, three knots were tied outside of the simulator box and a knot pusher was used to slide and tie the knot at the suture site.

Experimental Design:

Each participant performed three trials of the 13 tasks on the MIST-VR, as per the basic module design provided by the system designers, and one trial of the five tasks on the MISTELS, as described by its designers. The order of trainers was counterbalanced. The order of the tasks within each trainer performance followed the order as listed above. We did not change the tasks, even though they are different, as there are no directly comparable tasks between the two simulators.

Dependent Measures:

Overall scores for the MIST-VR system are based on the sum of time to task completion, errors, and economy of movement, all of which are computed by the system software that tracks the user's input through the tools. Economy of movement is calculated from the deviations of tool movements from a predetermined optimal trajectory. A MISTELS score follows a normalized standard scoring system established by the creator of the system, Dr. Gerald Fried of McGill University. This score is also based on time for task completion and number of errors committed, but does not include efficiency of motion. An overall score with weighted averages for all five tasks was calculated for each participant. This study specifically compared time and scores.

Analysis:

For analysis, participants were divided into three groups by level of surgical training: medical students and interns (juniors), second and third year residents (midlevels), and fourth year residents and fellows (seniors). An analysis of variance (ANOVA) was done with an alpha value of 0.05 on the time and scores of both simulators. Only significant results are reported in this paper.

RESULTS AND DISCUSSION

Not all tasks on the MIST-VR were able to distinguish between experience levels in terms of time, error, and scores. Time to task completion was significantly different as a function of experience for eight of the 13 tasks (see Table 1). Post-hoc tests showed that for all of the tasks, with the exception of Withdraw Insert, the difference was found between the juniors and the other two groups of surgeons. Figure 1 is an example of the trend illustrated by the majority of tasks. For the task of Withdraw Insert, the difference lies between the senior surgeons and the other two groups.

Table 1: Statistics for MIST-VR task times.

Task	F	df	*p*
Acquire Place	3.337	2	0.042
Traversal	3.83	2	0.027
Withdraw Insert	4.56	2	0.014
Diathermy	3.64	2	0.032
Manipulate Diathermy	4.63	2	0.013
SC Clip	4.05	2	0.022
SD Diathermy	5.39	2	0.007
Start Stitch	4.67	2	0.013

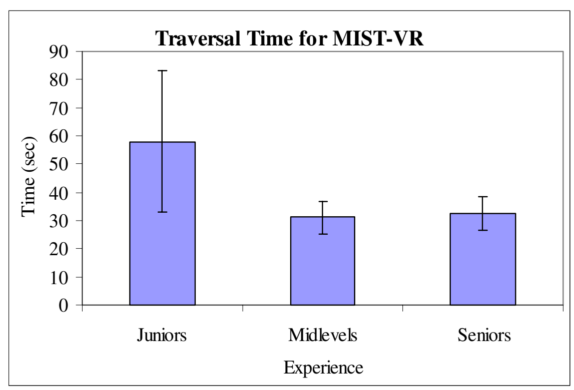

Figure 1: Traversal time to completion versus experience level for the MIST-VR.

The scores from the MIST-VR were calculated from the addition of the time to task completion, number of errors, and efficiency of motion. The lower the score, the better the performance. Scores on five of the tasks showed a statistically significant difference between skill levels. Post-hoc analyses of these tasks revealed that the difference was between the junior surgeons and the more advanced surgeons, except in the Start Stitch task, which showed an improvement in score from the less practiced surgeons (junior and midlevels) to the senior group (see Figure 2).

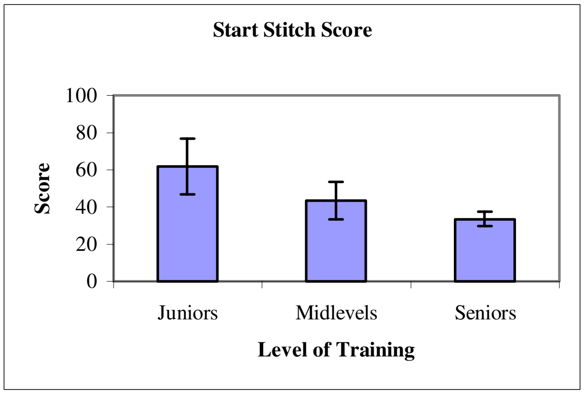

Figure 2: Score in Start Stitch task versus level of experience for the MIST-VR.

The MISTELS system was able to differentiate between experience levels on three of the five tasks in time, and four out of five tasks in score (see Table 2). In general, the more experienced the surgeon, the better the time and score were for the task. In this case, the higher the score, the better the performance. The senior surgeons performed better than the midlevels, who did better than the juniors for all tasks showing significant differences, in both time and scores. The overall score demonstrated a significant increase with experience ($F_{(2,20)}$=11.72, $p \leq 0.001$). The overall score is the combined weighted scores of the five individual tasks. Figure 3 illustrates an example of the trend found for all tasks showing significant differences.

Table 2: Statistics for MISTELS task scores.

Task	F	df	*p*
Peg Transfer	7.55	2	0.003
Pattern Cutting	8.85	2	0.001
Endoloop	4.69	2	0.020
Extra-corporeal Knot	6.23	2	0.007

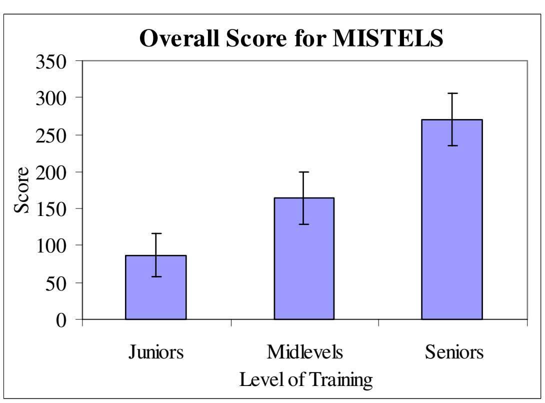

Figure 3: Total score versus experience level for the MISTELS.

Our results support previous studies that have shown that the MISTELS can record better performance for more experienced surgeons (Derossis et al, 1998; Derossis, Antoniuk, & Fried, 1999; Keyser, Derossis, Antoniuk, & Sigman 2000). Similarly, our results support those studies that showed the MIST-VR to give better scores for surgeons with higher experience (Chaudhry, Sutton, Wood, Stone, & McCloy, 1999; Gallagher, Richie, McClure, & McGuigan, 2001; Hamilton, Scott, Fleming, Rege, Laycock, Bergen, Tesfay, & Jones, 2002; Taffinder, Sutton, Fishwick, McManus, & Darzi, 1998). However, our comparison shows that the physical system was more capable of distinguishing between experience levels than the virtual system. The physical trainer discriminated between all three skill levels within the most of the tasks. It also showed that the more experience the participant had, the better and faster they performed. For those tasks that did not show a difference in performance as a function of experience, they were either too easy or too difficult. In particular, the Intra-corporeal knot-tying task is a fairly difficult task and is not usually taught and mastered until the experienced level which may have produced a ceiling effect in our subject pool.

The virtual system found a distinction only between the juniors and the other two experience levels and this only on certain tasks. With the exception of the Manipulate Diathermy task, all of the tasks where participants showed improved performance with experience were single skill tasks. The multiple skill tasks require a combination of two or more basic skills and were not able to differentiate between skill levels. A possible explanation for this result may hide behind the limited number of participants in each experience level. Furthermore, even though the tasks and performance metrics were designed specifically for each system, it appears that the performance metrics may be key to their usefulness. The MIST-VR scores were calculated by the addition of time to completion, errors committed, and economy of motion. The MISTELS scores, on the other hand, were weightened by the importance of the skill to laparoscopic surgery and the theoretical detriment of the errors.

It appears that the physical box trainer may be more representative of real motor tasks with the expected kinesthetic and haptic feedback from the task space, and thus was able to better reflect the skill level of the subject. Since virtual reality systems are a new concept, it is possible that the participants were less familiar with it than with a physical model. Virtual reality simulators require a translation from the virtual image to fine motor skills in the real world. The virtual reality trainer also lacks haptic feedback which possibly leads to a greater gap in familiarity. The differences in performance as a function of experience may not be measurable until residents have become familiar with the new technology.

CONCLUSION

Our research suggests that real simulators may presently be a more effective training tool for residency programs, as they are shown to better reflect the progress of acquisition of laparoscopic skills. However, as virtual reality technology becomes more popular and better developed, we may see a change in its value and usefulness. Further studies are needed in order to evaluate the ability of the physical and virtual reality simulators to distinguish between the various levels of experience and their effectiveness to train skills that are transferable to the OR setting.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under Grant # 0238284. We also acknowledge the participation of the residents at Tufts-New England Medical Center and Stamford Hospital, and the students of Tufts School of Medicine.

REFERENCES

Chaundhry, A., Sutton, C., Wood, J., Stone, R., McCloy, R. (1999). Learning rate for laparoscopic surgical skills on MIST-VR, a virtual reality simulator: quality of human-computer interface. *Annals of the Royal College of Surgeons in England* 81:281-286.

Derossis, A.M., Fried, G.M., Abrahamowicz, M., Sigman, H.H., Barkun, J.S., Meakins, J.L. (1998). Development of a model for training and evaluation of laparoscopic skills. *American Journal of Surgery* 175(6): 482-487.

Derossis, A.M., Antoniuk, M., Fried, G.M. (1999). Evaluation of laparoscopic skills: a 2-year follow-up during residency training. *Canadian Journal of Surgery* 42:293-296.

Fitts, P.M (1964). Perceptual-motor skills learning. In A.W. Melton (Ed.), *Categories of human learning* (pp. 243-285). New York: Academic Press.

Fried, G.M., Derossis, A.M., Bothwell, J., Sigman, H.H. (1999). Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. *Surgical Endoscopy* 13(11): 1077-1081.

Gallagher, A.G., Richie, K., McClure, N., McGuigan, J. (2001). Objective psychomotor assessment of experienced, junior and novice laparoscopists with virtual reality. *World Journal of Surgery* 25(11): 1478-1483.

Hamilton, E.C., Scott, D.J., Fleming, K.B., Rege, R.V., Laycock, R., Bergen, P.C., Tesfay, S.T., Jones, D.B. (2002). Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surgical Endoscopy* 16:406-411.

Hoffman, H., Dzung, V. (1997). Virtual reality: taching tool of the twenty-first century? *Academic Medicine* 12: 1076-1081.

Keyser, E.J., Derossis, A.M., Antoniuk, M., Sigman, M.M. (2000). A simplified simulator for the training and evaluation of laparoscopic skills. *Surgical Endoscopy* 14:149-153.

Munz, Y., Kumar, B.D., Moorthy, K., Bann, S., Darzi, A. (2004). Laparoscopic virtual reality and box trainers: Is one superior to the other? *Surgical Endoscopy* 18: 485-494.

Peters, J.H., Fried, G.M., Swanstrom, L.L., Soper, N.J., Sillin, L.F., Schirmer, B., Hoffman, K., SAGES FLS Commitee (2004). Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 135:21-27.

Schmidt, R.A. (1998). The processes of learning. In: *Motor control and learning: a behavioral emphasis* (pp. 457-491). Illinois: Human Kinetics Publishers, Inc.

Seymour, N.E., Gallagher, A.G., Roman, S.A., O'Brien, M.K., Bansal, V.K., Andersen, D.K., Satava, R.M. (2002). Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Annals of Surgery* 236(4): 458-464.

Taffinder, N., Sutton, C., Fishwick, R.J., McManus, I.C., Darzi, A. (1998). Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgey: results from randomized controlled studies using the MIST VR laparoscopic simulator. *Student Health and Technology Information* 50:124-130.