

Assessing Skill and Learning in Surgeons and Medical Students Using a Force Feedback Surgical Simulator

O'Toole, R.^{1,2}, Playter, R.¹, Krummel, T.³, Blank, W.¹, Cornelius, N.¹, Roberts, W.¹, Bell, W.¹, and Raibert, M.¹

¹ Boston Dynamics Inc., Cambridge, MA, USA.

² Harvard Medical School, Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA.

³ Department of Surgery, Hershey Medical Center, Penn State University, Hershey, PA, USA.

Abstract. We have developed an interactive virtual reality (VR) surgical simulator for the training and assessment of suturing technique in the context of end-to-end anastomosis. The surgical simulator is comprised of surgical tools with force feedback, a 3D visual display of the simulated surgical field, physics-based computer simulations of the tissues and tools, and software to measure and evaluate the trainee's performance. This study uses the simulator to compare the skills of experienced vascular surgeons to medical students. Eight parameters were measured to evaluate performance during VR suturing tasks. The data indicate significant differences between surgeon and non-surgeon performance, as well as improvement in performance with training. We believe that this study offers support for the use of virtual reality surgical simulators to augment surgical skill assessment and training.

Introduction

Although surgical technique has evolved considerably over time, the process of training surgeons has undergone little change since the inception of the Halstedian technique over a century ago [barnes89]. Surgery is in many ways a traditional hands-on apprenticeship. Surgical residents acquire skill by first observing experienced surgeons in action, and then performing progressively more of the surgical procedures themselves as their training progresses. As their skill levels increase, the residents are given increasing responsibility. The assessment of the trainee's skill is performed subjectively by senior surgeons.

We are exploring the use of interactive virtual reality computer simulations to augment training on patients. Virtual reality (VR) refers to a computer-simulated environment that provides sensory output to the user in an attempt to mimic a real environment. A particularly successful example of virtual reality training is the use of aircraft simulators for the training of commercial and military pilots [Higgins97]. With the exception of one flight in a real aircraft, commercial pilots regularly do all

of their training to upgrade from one aircraft to another on simulators. The potential advantages of VR to augment training include:

- **Quantification** of performance and progress for training and accreditation.
- **Standardization** of training regimens independent of patient population.
- **Exposure** to rare but important situations in an era of shrinking training opportunities.
- **Reduction of risk** to patients.
- **Reduction of cost** through more effective use of operating room time.
- **Improved** educational techniques.

The potential benefits of VR training have motivated the development of VR surgical simulators during the past several years [satava96]. These simulators have advanced from fly-through applications [satava93] to more recent work incorporating real surgical tools and force feedback [baumann96, fischer95, singh94, mcdonald95]. Research groups are now beginning to attack the problem of validating surgical simulators [Weghorst98, Taffinder98].

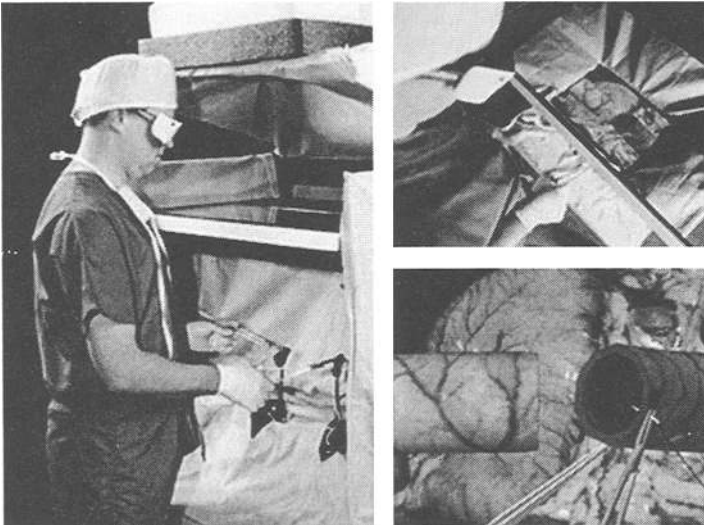


Fig. 1. The BDI surgical simulator lets you see, touch, and feel simulated organs using 3D computer graphics, physics-based simulations, and advanced force feedback. Using a needle holder and forceps attached to force feedback devices, the user can grasp, poke, pluck, and suture flexible tube organs. The user can feel the vessels when touched and the vessels move realistically in response to touch. The surgical station uses a mirror arrangement to place the 3D image of the patient in the correct position relative to the user. This system allows the user to control a virtual needle and thread to perform a simulated end-to-end anastomosis.

We have developed a surgical simulator for the evaluation and training of surgical skills for anastomosis [Playter 1997]. Anastomosis was chosen as a first application as it is a common surgical task that has a high reliance on proper technique. The

simulator is an integrated system that allows users to perform a simulated end-to-end anastomosis using a virtual needle and thread [Figure 1]. The simulator is specialized for touch and interactivity. Users feel virtual organs when they touch, probe, grasp, and suture simulated biological tissues. Physics-based simulation is used to make the flexible vessels feel realistic and be responsive to touch. Instrumented surgical tools mounted to force-feedback devices act as two-way connections to the simulated world; they measure the position and configuration of the tools and apply interaction forces to the user's hands. The user's view of the surgical field is created using real-time, 3D computer graphic images of bodily organs, tissues, and surgical tools. The visual image is projected into a natural position relative to the surgeon using a mirror system. The anastomosis simulator is able to measure many aspects of the user's performance including the forces applied to the tissues as well as the actual motions of the tools in space. By recording these values we can quantify the user's performance. A force history plot of the forces exerted by the needle, as well as 3D needle guide vectors to indicate ideal needle orientation, can be displayed during the simulation to guide training. A digital video is automatically created during each training session so that the user may replay performances.

In this paper we describe a preliminary study aimed at investigating two key questions regarding surgical simulation:

1. Can surgical simulation be used to *measure* surgical skill?
2. Can surgical simulation be used to *train* surgical skill?

To address these issues we simulated the essential elements of suturing technique in the context of end-to-end anastomosis and devised metrics for evaluating performances. We then measured and compared the performances of subjects of presumably different surgical skill levels. The skilled and unskilled users of our surgical simulator were experienced vascular surgeons and medical students, respectively. Our hypothesis was that if the trainer could measure surgical skill, then these two groups should have different levels of performance on the simulator. Further, if the trainer can be used to improve surgical skill, then the performances should improve over time, particularly amongst the medical students.

Methods

In an assessment of our surgical simulator we designed and implemented a human performance study to test if the surgical simulator can measure or train surgical skill. We chose suturing on a large flexible vessel as the surgical task. We used a simplified version of the surgical simulator that included only a needle driver, a curved needle and one flexible vessel with four small targets near the edge. The goal of the surgical task was to 'suture' the vessel in sequence at the targeted locations. A 'suture' consisted of passing the needle completely through the vessel. The completion of four sutures is defined as one trial [Figure 2].

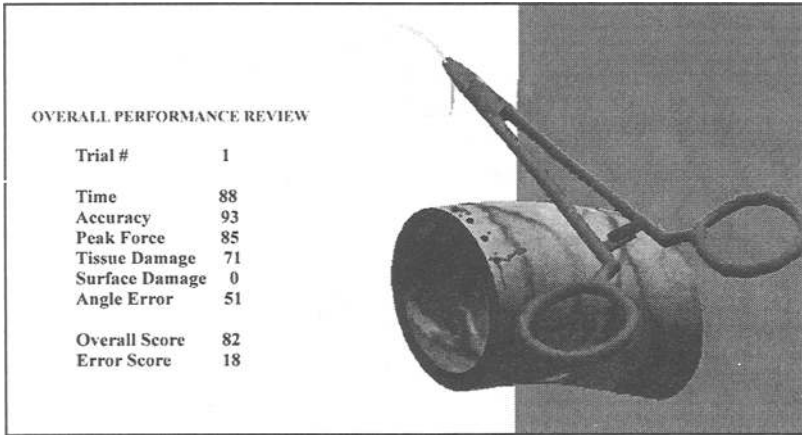


Fig. 2. A performance evaluation is displayed to the user at the completion of each trial. The Overall Score equals 100 minus the Error Score. The four goal target locations for the surgical task are shown on the edge of the vessel. Passing the needle completely through the tissue four times is considered one trial. Note the four goal targets the scratch marks on the surface of the vessel near the target locations.

Seven parameters were used to evaluate performance during the simulated suturing task: total tissue damage, peak force to the tissue, accuracy of stitches, time to complete the task, damage to the surface of the tissue, angular error in needle technique, and an overall error score. The scoring in each category, as well as an overall error score, is displayed to the user at the completion of each trial [Figure 2]. An eighth parameter, the total distance traveled by the tip of the tool, was also measured, but the value was neither displayed to the users at the completion of the task, nor included in the overall error score.

The participants were eight currently practicing vascular surgeons from four Boston hospitals and twelve medical students from Harvard Medical School. The eight surgeons ranged in experience from 7 to 25 years in practice (mean 14.5 years), and had performed between 1100 and 7000 vascular surgeries (mean 3250 surgeries). The surgeons were all right handed, were all male, and had an average age of 43.5 years (range from 31 to 56). The medical students were first, second and third year medical students from Harvard Medical School (three first years, six second years, and three third years). Eleven of the twelve were right handed, and eight of the twelve were male. Two of the third years had already completed their core surgery rotation, but otherwise none of the students had any surgical experience. The medical students' average age was 24.6 years (range of 20 to 30). None of the surgeons or students were involved in the design of the system or associated with our research group, and none had previously used the system. Each participant in the study performed an identical protocol, and was tested in isolation from other test subjects. After collecting demographic data on each subject and a brief initial training period with a non-surgical VR application, the experimental protocol was performed. All experiments were performed between February and March of 1997. The protocol consisted of each user performing the previously described VR task (4 'sutures' into a

single large vessel) 15 consecutive times, under varying test conditions. These conditions included using the dominant hand (test condition 1), using the non-dominant hand (test condition 2), and using the dominant hand with some additional visual guidance (test condition 3). For all analyses, statistical significance was defined as a p value ≤ 0.05 using two-sided Student's t test.

Results

Comparison of Surgeon and Medical Student Performances

For experimental condition 1 (dominant hand), the surgeons outperformed the students in all 8 parameters, and 6 of these differences were significant [Figure 3]. There were statistically significant differences, with the average medical student performance being worse than the average surgeon performance, for time, tissue damage, angle error, tip distance, accuracy error, and overall error. Similar results were obtained for the other test conditions, but these data are not included in this text.

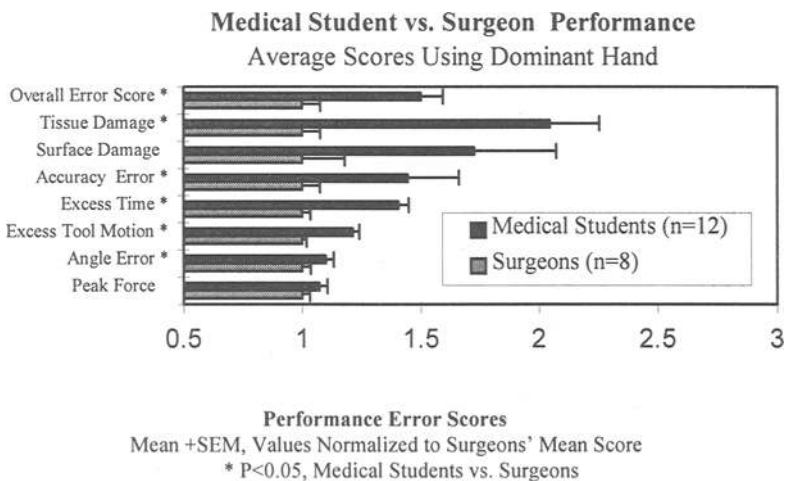


Fig. 3. The bar graph compares the average surgeon and medical student performances for the 8 measured parameters. The average values have been normalized to the surgeons' average score. Larger values indicate worse performance. The surgeons average was better than the medical students average for all 8 parameters. Six of these differences were statistically significant at the $p < 0.05$ level. This data is for Test Condition 1 where the participants used their dominant hand for the task.

Changes in VR Performance with Training

Both the surgeons and medical students tended to improve during the training session as demonstrated by data for the metric tissue damage [Figure 4]. To determine the magnitude of the users' performance change with training, the scores from trials 1 and

2¹ were compared with trials 13 and 14. All four of these trials were performed under the same test condition (condition 1 - dominant hand). The only difference between these two sets of trials is that the user performed 10 training trials between the two sets. The data demonstrate that for seven of the eight parameters, both the surgeons and the medical student performances improved [Figure 8].

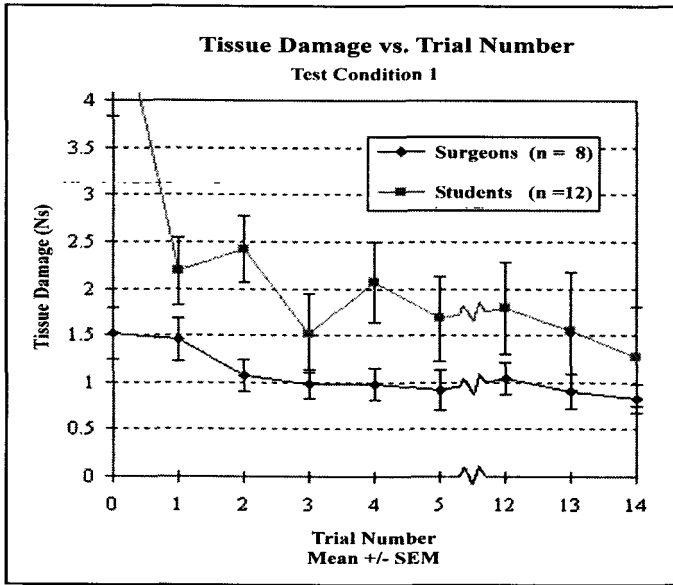


Fig. 4. The average Tissue Damage performance versus trial number is shown for trials 0 to 6, and 12 to 14, the trials under test condition 1. Larger values represent worse performance. Note that the students initially have a very poor performance and improve with time. The surgeons also improve with time, but to a lesser extent as quantified in Figure 5.

The magnitude of performance improvement was larger for the students than the surgeons in six of the eight measured parameters [Figure 5]. In five of these cases these differences were statistically significant. Depending on the parameter, the improvements were of differing magnitudes. For example there was a large improvement by the medical students for the Surface Damage parameter and a very small improvement in the Tip Distance parameter. Only for the case of the surgeons' Angle Error does the average score actually become worse during the training session.

¹ Trial 0 was not included in the Before Training data because trial 0 is slightly different than the other trials as the users are not yet familiar with the system. Including trial 0 would only strengthen our conclusions, as the medical students performed poorly in trial 0. Excluding trial 0 is therefore the conservative approach.

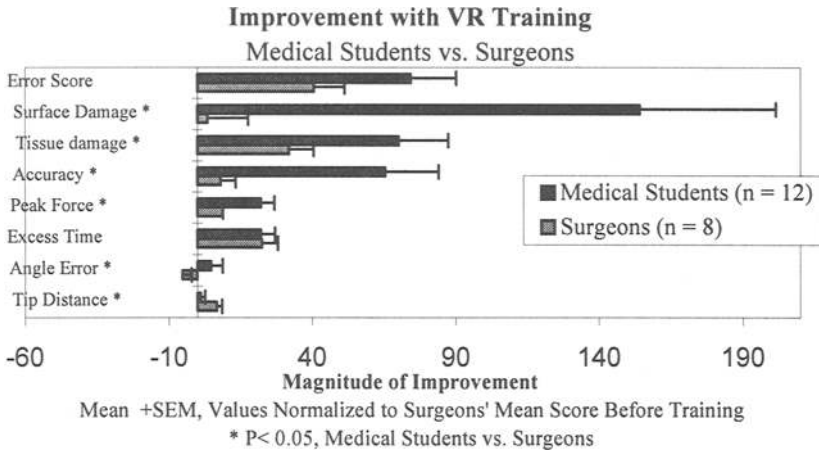


Fig. 5. The bar graph compares the average magnitude of improvement for the surgeons and medical students. Larger values indicate larger amounts of improvement. The values are the raw magnitude of improvement divided by the surgeons' average value before training. The medical students had statistically significant larger improvements for 5 of the cases, indicating that their scores improved more than the surgeons' scores during the training session. The surgeons had a larger improvement for Tip Distance, the only parameter for which the participants were not given feedback concerning their performance during the training sessions.

Discussion

This study aims to test if the surgical simulator can measure or train some component of surgical skill. Assuming that the surgeons and medical students have different levels of skill, one would expect their performances to differ if the simulator measured some component of surgical skill. If the surgical simulator was able to train some component of surgical skill, one would expect that the performance would improve with time and that the medical students would improve more than the surgeons.

There were statistically significant differences between the surgeons' and medical students' average performances on 6 of the 8 test cases. This data demonstrates difference between the two groups, with the surgeons generally outperforming the students. The magnitude of the difference between the two groups varied depending upon the parameter. For example, there were large differences in the two groups in the tissue damage scores, whereas the peak force scores tended to be much more similar. Parameters with large differences between the medical students and surgeons may represent more useful indicators of surgical skill. Tissue damage, excess time, excess tool tip motion, and overall error score all seem to be useful metrics, since there were statistically significant differences between the two groups for all three test conditions.

There are several possible explanations for the cases where there were no significant differences between the two groups. The most obvious explanation is that

the present study lacks statistical power to observe these differences. A study with a larger sample of medical students and surgeons might yield stronger results. A second idea is that perhaps some of these values represent quantities that are not different between skilled and unskilled surgeons, and should not be included in the performance evaluations. There is no previous evidence that the metrics used in this study to evaluate performance are valid metrics of surgical skill so this is certainly plausible that these are not appropriate metrics. And finally, the precise implementation of the skills assessment may not be able to capture differences between the two groups. For example, the goal value for the accuracy parameter may have been too easy to achieve for both groups. If the goal value was too easy, neither the surgeons nor students had incentive to perform with maximum accuracy, so differences in accuracy skill may not be demonstrated.

Is it possible that the observed performance dichotomy between the two groups was the basis of some difference between the two groups besides surgical skill? There are several factors that differed between the surgeons and the medical students: age (the medical students were younger), gender (33% of the medical students were women, none of the surgeons were), and having chosen surgery as a career (all of the surgeons chose surgery as a career, only about a fifth of the students would be expected to choose surgery).

Each of these three factors will be considered individually. First, there was a rather large difference in age between the two groups; however, one would not expect the older group to perform better in a manual dexterity task. In fact, there is evidence that manual dexterity skills tend to deteriorate with age [Scheuneman89]. Therefore one would expect the younger medical students to do better, not worse, than the surgeons if the performance differences were based upon age. Second, there were more women in the medical student group. As with age, there is no clear evidence that women would tend to do worse on a manual dexterity task. Again, there is evidence that women tend to do better [Harris94, Scheuneman89], on some components of manual dexterity tasks, so one might expect the gender difference to improve medical student performance, not hinder it.

Finally one might theorize that medical students who choose surgery are self-selected for greater manual dexterity or visuospatial ability than their colleagues who do not choose surgery. Presumably the medical student group in this study contains a large percentage of students who will not become surgeons. However there is evidence debunking the idea that medical students who choose surgery have better manual dexterity or visuospatial skills [Harris94, Squire89]. No study has yet shown differences in general manual dexterity between medical students who chose surgery and those who do not. A task must be very similar to a specific surgical skill for the surgeons to outperform the non-surgeons. It is therefore not likely that either age, gender, or choice of medical career path explain our results.

In addition to the observed differences in performance between the two groups, the data also demonstrate changes in performance with repeated training. During the course of the training session, the medical student average performance improved in all 8 measured categories. Four of the parameters showed statistically significant improvements (time, peak force, tissue damage, and error score), and two more were

very close to statistical significance (surface damage ($p = 0.06$) and accuracy ($p=0.07$)). The other two values, tip distance and angle error were not significantly improved and had the smallest magnitude of improvement. The fact that tip distance did not improve is not surprising since this is the only parameter about which the participants were given no feedback during the experiment. Without information regarding this parameter they had no incentive to improve the score in this area.

In all of the categories except time and tip distance, the medical students average improved by a larger magnitude than the surgeons. The surgeon average did improve by a statistically significant margin for 4 of the 8 parameters (time, tip distance, tissue damage, and error score). Some improvement in performance is expected as the trainees become accustomed to the test apparatus and the surgical task, so it is not surprising that both the surgeons and medical students showed improvement. However, the larger, statistically significant improvement demonstrated by the students in 5 of the 7 parameters in the surgical score (recall that tip distance was not a component of the overall error score) indicates that the medical student scores improved more during the training session. This result is consistent with the idea that if the surgical trainer can train surgical skill, the medical students would benefit more from a training session than the surgeons.

One explanation for the larger improvements during the trials by the students is that the students were acquiring surgical skills that the surgeons already had. An alternate explanation is that the students were simply faster learners than the surgeons. Whatever the reason, one should be cautious about extrapolating these results to imply that the medical students were learning surgical skill. Since there was no long term follow-up on this single training session, there is no evidence that any of the improvements in performance were permanent. And further, there is not yet evidence that the observed improvements correlate with improvements in surgical skill. Be that as it may, the data are a first step in validating the use of surgical simulators for training of surgical skill.

Conclusions

To our knowledge we have created the first 3D VR surgical simulator to measure suturing skills and have used it to demonstrate significant performance differences between medical students and surgeons. Our VR surgical trainer is unique because it provides quantification and evaluation of a user's performance, and focuses on the realistic sense of touch by allowing users to reach into the virtual environment with tools to touch and manipulate simulated tissues.

Our current research is aimed at answering two broad questions: 1) Can we *measure* surgical skill using VR technology? and 2) Can we *train* surgical skill using VR technology?

Strictly speaking, we have not yet addressed either of these questions directly. To directly answer the first question we must assess surgical skill using some other traditional method of evaluating skill and determine if VR surgical skill is correlated to real surgical skill. To fully answer the second question, we must compare actual surgical skill with and without VR surgical training over time. In this initial study we

have avoided the issue of evaluating real surgical skill. Instead we have focused on evaluating skill level and learning using only VR technology.

Our initial study compares the performance of two groups of subjects that are assumed to have different levels of real surgical skill: practicing vascular surgeons and medical students. The performances of these two groups were distinguishable from one another, suggesting that the simulator could be measuring components of surgical skill. Similarly, improvements in performance with practice suggests that the simulator may be capable of teaching components of surgical skill. Future clinical validation studies will be needed to verify that the trainer can be used to measure and train actual surgical skill. However, the current study does provide data indicating that VR technology may play an important role in training tomorrow's surgeons

Acknowledgments

The surgical simulator has been developed as part of DARPA's Advanced Biomedical Technologies Program. We are indebted to Dr. William Quist from the Beth Israel-Deaconess Medical Center who provided significant clinical input and assistance to the project. Many thanks are also due to the Harvard Medical School students and Boston area vascular surgeons who generously gave their time to participate in the study.

References

- Barnes, R. W., Lange, N.P., Whiteside, M.F.],Halstedian Technique Revisited, *Annals of Surgery*, 210, 1989: 118-121.
- Barnes, R. W.,Surgical Handicraft: Teaching and Learning Surgical Skills, *The American Journal of Surgery* 153, 1987: 422-427.
- Playter, R., Raibert, M.,A Virtual Surgery Simulator Using Advanced Haptic Feedback, *Journal of Minimally Invasive Therapy*, To Appear, 1997.
- Satava, R. M., Medical Virtual Reality, *The Current Status of the Future*. S. Weghorst, H. Siegurg and K. Morgan Eds., *Health Care in the Information Age*. IOS Press, Amsterdam, 1996: 542-45.
- Satava, R. M., Virtual Reality Surgical Simulator: The First Steps, *Surg Endosc* 7, 1993: 203-05.
- Reinig K., C. Rush, H. Pelster, V. Spitzer, and J. Heath, Real-Time Visually and Haptically Accurate Surgical Simulation. S. Weghorst, H. Siegurg and K. Morgan Eds., *Health Care in the Information Age*. IOS Press, Amsterdam, 1996: 542-45.
- Baumann, R., D. Glauser, D. Tappy, C. Bauer, and R. Clavel, Force Feedback for Virtual Reality Based Minimally Invasive Surgery Simulator. S. Weghorst, H. Siegurg and K. Morgan Eds., *Health Care in the Information Age*. IOS Press, Amsterdam, 1996: 564-579.
- Fischer, H., B. Neisius and R. Trapp, Tactile Feedback for Endoscopic Surgery. K. Morgan, R. Satava, H. Sieburg, R. Matheus, and J. Christensen Eds., *Interactive Technology and the New Paradigm for Healthcare*, IOS Press, Amsterdam, 1995: 114-117.
- Singh, S., M. Bostrom, D. Popa, and C. Wiley, Design of an Interactive Lubar Puncture Simulator with Tactile Feedback. *Proceedings of IEEE International Conference on Robotics and Automoaion*, IEEE, New York, 1994: 1734-1752.

- McDonald, J., L. Rosenberg and D. Stredney, *Virtual Reality Technology Applied to Anesthesiology*. K. Morgan, R. Satava, H. Sieburg, R. Mattheus, and J. Christensen Eds., *Interactive Technology and the New Paradigm for Healthcare*, IOS Press, Amsterdam, 1995: 237-243.
- Steele, R. J. C., Walder, C, and Herbert, M., Psychomotor testing and the ability to perform an anastomosis in junior surgical trainees, *Br. J. Surg*, 1992, **79**:1065-1066.
- Harris, C.J., Herbert, M., and Steele, R.J.C., Psychomotor skills of surgical trainees compared with those of different medical specialists, *Br. J. Surg*, 1994, **81**: 382-283.
- Squire, D., Giachino, A.A., Profit, A.W., Heaney, C., Objective comparison of manual dexterity in physicians and surgeons *Can J Surg* 1989; **32**:467-76.
- Scheueneman, A., Pickleman, J., Freeark, R. Age, gender, lateral dominance, and prediction of operative skill among surgical residents. *Surgery* 1983; **98**: 506-13.
- Higgins, G. A., Merrill, G.L., Hettinger, L.J., Kaufman, C.R., Champion, H.R., and Satava, R.M. *New Simulation Technologies for Surgical Training and Certification: Current Status and Future Projections*, *Presence*, 6 (2) April 1997, 160-172.
- Weghorst, S., Airola, C., Oppenheimer, P., Edmond, C.V., Patience, T., Heskamp, D., and Miller, J., Validation of the Madigan ESS Simulator, In: Westwood, J.D., Hoffman, H.M, Stredney, D. And Weghorst, S.J. *Medicine Meets Virtual Reality: Art, Science, Technology: Healthcare (R)Evolution*, IOS Press, 1998, pg.399-405.
- Taffinder, N., Sutton, C, Fishwick, R.J., McManus, I.C., and Darzi, A. Validation of Virtual Reality to Teach and Access Psychomotor Skills in Laparoscopic Surgery: Results from Randomised Controlled Studeies Using the MIST VR Laparoscopic Simulator, In: Westwood, J.D., Hoffman, H.M, Stredney, D. And Weghorst, S.J. *Medicine Meets Virtual Reality: Art, Science, Technology: Healthcare (R)Evolution*, IOS Press, 1998, pg.124-130.