

# Virtual Reality Applied to Procedural Testing: The Next Era

Matthew B. Bloom, MD, MS, Chantal L. Rawn, BS, Arnold D. Salzberg, MD, and Thomas M. Krummel, MD, FACS

*From the Center for Advanced Technology in Surgery, Department of Surgery, Stanford University School of Medicine, Stanford, California*

---

## Objective

To establish the construct validity of a virtual reality-based upper gastrointestinal endoscopy simulator as a tool for the skills training of residents.

## Summary Background Data

Previous studies have demonstrated the relevance of virtual reality training as an adjunct to traditional operating room learning for residents. The use of specific task trainers, which have the ability to objectively analyze and track user performance, has been shown to demonstrate improvements in performance over time. Using this off-line technology can lessen the financial and ethical concerns of using operative time to teach basic skills.

## Methods

Thirty-five residents and fellows from General Surgery and Gastrointestinal Medicine were recruited for this study. Their performance on virtual reality upper endoscopy tasks was analyzed by computer. Assessments were made on param-

eters such as time needed to finish the examination, completeness of the examination, and number of wall collisions. Subjective experiences were queried through questionnaires. Users were grouped according to their prior level of experience performing endoscopy.

## Results

Construct validation of this simulator was demonstrated. Performance on visualization and biopsy tasks varied directly with the subjects' prior experience level. Subjective responses indicated that novice and intermediate users felt the simulation to be a useful experience, and that they would use the equipment in their off time if it were available.

## Conclusions

Virtual reality simulation may be a useful adjunct to traditional operating room experiences. Construct validity testing demonstrates the efficacy of this device. Similar objective methods of skills evaluation may be useful as part of a residency skills curriculum and as a means of procedural skills testing.

---

Over the past 15 years, advances in surgical instrumentation and electronics have transformed the practice of modern surgery and allowed for the rapid acceptance of minimally invasive techniques. The development of computer-assisted technologies as training instruments will likely revolutionize the process of surgical education in a similar fashion.

The modern surgical residency traces its roots directly to the residency program at Johns Hopkins established by William Halsted at the turn of the century. "Learning by

doing" has since been the method through which surgeons in training have acquired their skills.<sup>1</sup> A predicate to this approach, however, is the availability of ample clinical material, as well as clinical instruction. The quality of training is influenced by chance patient encounters and subjective methods of skills assessment.

Much has been written about the financial and ethical considerations of allowing residents to practice their skills directly on patients. Traditionally, the operating room (OR) has been the classroom for surgical education, but the cost of OR time is ever-increasing.<sup>2</sup> It has been estimated that the annual costs to our healthcare system of OR time alone for training of chief residents exceeds \$50 million per year.<sup>3</sup> Increases in healthcare costs, coupled with decreasing Medicare financing of medical education, threatens funding for residents.<sup>4</sup> As OR time becomes increasingly precious,

---

The authors received no compensation, financial or otherwise, from 5DT Inc.

Correspondence: Thomas M. Krummel, MD, FACS, Department of Surgery, Stanford University School of Medicine, 701B Welch Rd., Suite 225, Stanford CA 94305-5784.

E-mail: tkrummel@stanford.edu

Accepted for publication August 7, 2002.

residency programs are forced to strike a balance between their business and academic responsibilities.

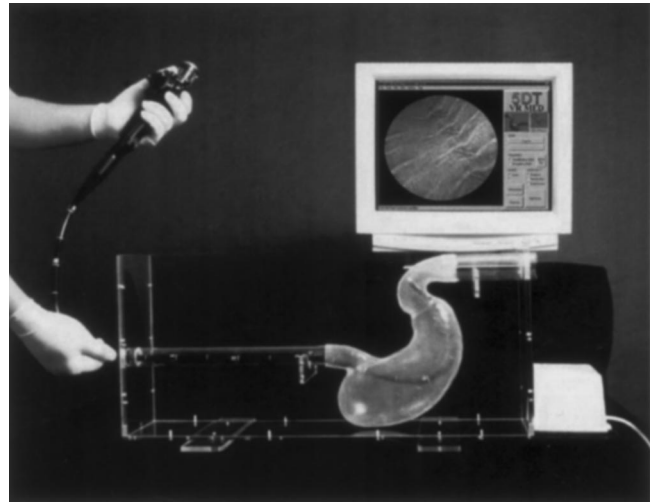
To address these problems, alternative methods of training, such as using cadavers or live animals, have been suggested. However, these solutions have been criticized as being expensive and unrealistic. Another alternative to assess and train young surgeons is the use of virtual reality (VR) technologies.<sup>5</sup>

The question has been raised: Can we measure procedural skills using VR-based methods, and can we further develop these skills in less experienced surgeons?<sup>6</sup> Supporters of VR and simulation have suggested that computer-based instruction can be an educational and ethical addition to the experience of training on live patients. It would afford residents practice on standardized teaching cases, as well as on rare and complicated procedures. Residents could review a case in preparation for the real OR and be presented with specific challenges drawn from a teaching library of real operations. Advances in computer graphics and rendering technologies offer simulations that are visually appealing and true to life. Other advances in tactile feedback devices allow for the sense of touch to further improve the perception of realism. Additionally, the computer-based systems allow for objective assessment of user performance and skills. Objective data regarding the motion of the instruments, accuracy, and forces can be acquired, quantified, and compiled into a performance report card.<sup>6,7</sup> Simulators may evolve into a standard method for resident evaluation.

Evaluating procedural proficiency is a difficult task. Several studies have demonstrated that psychomotor and dexterity skills testing alone does not correlate well with OR performance.<sup>8-11</sup> Other criteria such as stress tolerance and tests of visuospatial abilities have demonstrated better correlation.<sup>12-14</sup> Directly testing procedural skills through computer simulations may offer a more reliable measure of one's ability to perform specific tasks.<sup>15</sup>

For simulators to gain acceptance into training programs, they need to be evaluated for their efficacy in teaching and evaluating procedural skills. Early validation studies employing simulations of a hollow tube anastomosis and laparoscopic manipulation have demonstrated differences in performance criteria between novice and expert surgeons, improvements in performance over time, and a correlation between performance on the simulator and actual performance on patients.<sup>6,16</sup>

An essential step in the formal evaluation of any simulator is the demonstration of its construct validity: that is, verification that the simulator awards performance scores that correspond to the ability level of the individuals performing the actual task. More experienced users are expected to achieve better scores than novices if the simulator contains both an accurate means of scoring and an accurate simulation of the procedure. Other evaluation steps include studies of the usability (expert opinion of the simulator's performance) and demonstrations of the transferability of skills learned in a VR environment to the real world.



**Figure 1.** The Upper GI Endoscopy Simulator.

In this paper we report the validity testing of the Upper GI Endoscopy Simulator (5DT Inc., Santa Clara, CA). The Upper GI Simulator is one of several commercially available procedural simulators that simulate a specific task such as bronchoscopy, endoscopy, arthroscopy, intravenous insertion, and laparoscopy. Several of these other devices are being evaluated by our laboratory for their educational utility. This upper endoscopy device could be used by trainees in surgical and internal/family medicine specialties.

## MATERIALS AND METHODS

### Simulation Device

The 5DT Gastroscope Training Simulator (5DT Inc, Santa Clara, CA) is a multimedia device designed to simulate the critical steps of a routine upper endoscopy. A standard gastroscope is fitted with a tracking sensor at its tip that is used to track the tip's position and orientation. This information is fed into a standard PC (PIII 650MHz, 256MB RAM, GeForce256 video card) that hosts a 3D computer graphics model of the GI tract. As the instrumented gastroscope is inserted into a life-size transparent silicone rubber model of the esophagus, stomach, and duodenum, the computer renders an image that corresponds to what would be seen when using a gastroscope in a real patient (Fig. 1). The device provides a scorecard of the user's performance, based on internal scoring of items such as percentage of total surface area viewed, time to complete the task, number of wall collisions, and number of injuries caused by the biopsy tool. The instructor has the ability to design different case scenarios. This feature was used to design the current study.

### Study Design

Our study was designed to establish construct validity by including subjects with differing levels of endoscopy expe-

rience. Thirty-five subjects, consisting of junior and senior general surgery residents and GI medicine fellows from our institution, were asked to participate in this study. Each subject was scheduled for an appointment over a period of 6 weeks. Prior use of this simulator was the only criterion for exclusion. Each subject completed an identical study session consisting of pre- and post-simulator questionnaires, a video demonstrating the proper technique for manipulation of the scope, and several standard endoscopy tasks.

A pre-simulator questionnaire determined each subject's level of training and endoscopy experience. Subjects were also asked about their familiarity with computer applications and video games, and their expectations of simulation systems.

After the completion of the questionnaire, subjects watched a brief video "Controlling the Scope" (Gastroenterology Atlas of Images, 5DT Inc.) outlining basic endoscope dial control and handling. Subjects were then allowed 3 minutes to instrument the transparent mannequin with the gastroscope and directly see the effects of the controls on the motion of the tip. The computer screen displayed the enteric tract anatomy that would be visualized if the instrument were similarly positioned in a real patient. At the conclusion of 3 minutes, subjects were led to a lesion at the esophageal side of the simulated GE junction, directly in view, and instructed how to use the biopsy tool to sample the lesion.

For the remainder of the study, the transparent mannequin torso was covered with a heavy drape to mimic a real endoscopy procedure. Subjects were then asked to perform two separate tasks, a visualization task and a biopsy task. These tasks were chosen to highlight different skills. During the visualization task, subjects were instructed to inspect as much of the GI tract as they were able, in as short a time as possible, and were told that they would be scored on these two criteria. Timing was begun as the oropharynx was instrumented. The computer quantified the percentage of the GI tract visualized and timed the procedure. This task was chosen to highlight the subject's ability to manipulate the scope throughout the upper GI tract. In the biopsy section, subjects were instructed to find and biopsy a lone lesion located along the tract. They were informed that they were being scored on the length of time required to find and perform the biopsy, as well as the number of wall collisions with the instrument tip. This procedure emphasized the skillful manipulation and positioning of the endoscope, as well as the ability to coordinate the scope's motion with the biopsy tool. On three separate trials, each subject encountered a single lesion that was presented sequentially in the fundus, duodenum, and cardia. Subjects who could not find and biopsy the lesions after 10 minutes were told to concede. Subjects were not informed of the scores of their performances.

After completion of the endoscopy tasks, subjects were asked to complete a second survey. A Likert scale (1 to 5)

was employed to quantify the opinions of the subjects. This questionnaire explored their feelings about the usefulness of this device as a teaching tool and its realistic representation of gastroscopy.

## Statistical Methods

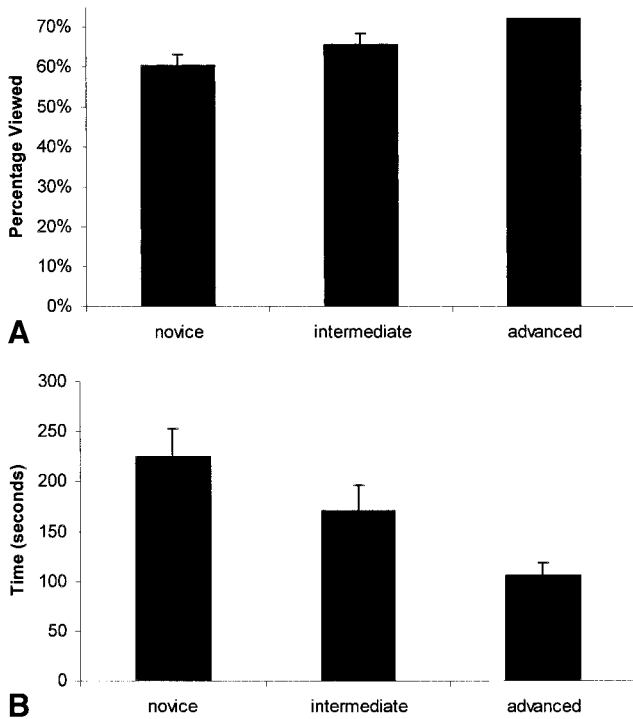
The data obtained from questionnaires and from the computer scoring of the subjects' performance of the tasks were entered into a database in a blinded fashion. Subjects were broken into skill level groups based on their prior level of endoscopic (both upper and lower) experience: none ( $n = 16$ ), 1 to 30 ( $n = 9$ ), and more than 30 endoscopies ( $n = 10$ ) performed in the past 5 years. A one-way ANOVA analysis of the data was performed to assess statistically significant differences between the groups. The mean differences and correlations were considered to be significant at the 0.05 level. All values represent means and are reported as mean  $\pm$  standard error. Questionnaire data were analyzed with nonparametric statistics.

## RESULTS

### Performance Scores

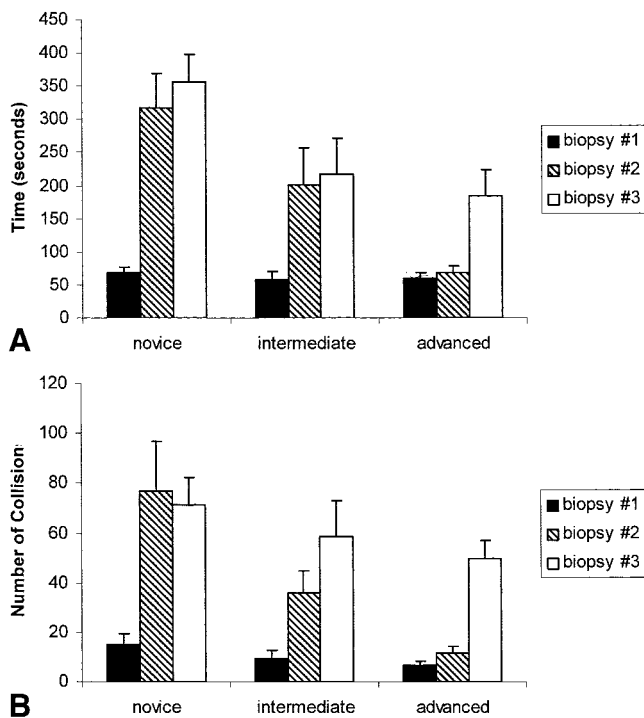
The visualization task was defined to test subjects' ability to navigate the scope through the upper GI tract. Subjects were instructed to visualize the GI tract as completely as possible, in as short a time as possible. The mean percentages of total surface area visualized by the groups were novice  $60.56 \pm 2.56$ , intermediate  $66.56 \pm 2.80$ , and advanced  $72.10 \pm 0.23$  (Fig. 2); differences between the groups were statistically significant ( $P = .005$ ). The average times (in seconds) required to complete this task were also statistically significantly different between the groups: novice  $224.81 \pm 27.65$ , intermediate  $171.22 \pm 25.43$ , and advanced  $106.40 \pm 13.08$  ( $P = .008$ ). The amount of prior endoscopy experience related directly to the percentage of the enteric tract that subjects were able to visualize, and inversely with the time required to complete their examinations.

As a separate task, subjects were instructed to locate and biopsy a distinct lesion in the GI tract. This task was designed to test the user's ability to coordinate both the scope and the biopsy tool. Biopsy 1 consisted of a lesion located in the fundus, directly in view on entering the stomach. In biopsy 1, no significant differences were seen among the groups under either criteria of time required to complete the task or number of wall collisions (Fig. 3). Biopsy 2 consisted of a lesion in the duodenum. This lesion obliged the user to navigate through the pylorus into the duodenum, requiring the technical ability of paradoxically moving the instrument tip cephalad from the greater to lesser curves of the stomach as the scope was advanced. In biopsy 2, a significant difference between groups was observed both in the time required to complete the task and the

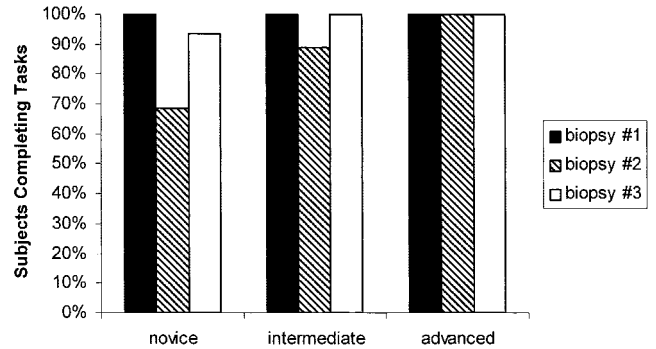


**Figure 2.** (A) Completeness of endoscopic visualization task. (B) Time to complete endoscopic visualization task.

number of wall collisions:  $316.31 \pm 53.81$ ,  $200.22 \pm 56.69$ ,  $68.70 \pm 9.71$  ( $P = .004$ ) and  $76.75 \pm 20.09$ ,  $36.22 \pm 8.60$ , and  $11.7 \pm 2.62$  ( $P = .022$ ). Direct relationships were seen



**Figure 3.** Time to complete biopsy tasks (A) and number of wall collisions during biopsy (B) versus experience level.



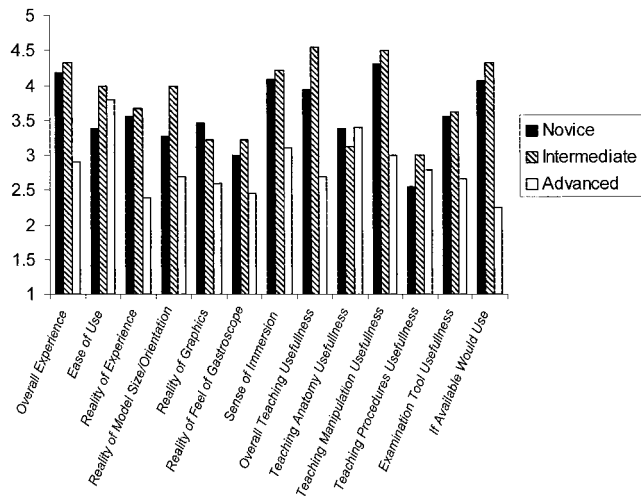
**Figure 4.** Percentage of subjects able to complete biopsy tasks.

between the level of experience and both quicker and safer performances. Biopsy 3 consisted of a lesion in the cardia. To find this lesion, users had to know to retroflex and withdraw the scope to look into the cardia from below. In biopsy 3, differences were demonstrated between the three groups in time to complete the task ( $356.00 \pm 42.60$ ,  $217.44 \pm 53.21$ ,  $183.90 \pm 39.88$ ;  $P = .020$ ) only. Although the number of collisions varied inversely with level of experience, differences between the groups were not significant ( $71.40 \pm 11.07$ ,  $58.78 \pm 13.95$ ,  $49.80 \pm 7.36$ ;  $P = .376$ ).

All advanced subjects were able to complete all the biopsy tasks in the time limit allowed, but only 11 of 16 novices and 8 of 9 intermediates were able to navigate into and biopsy the duodenal lesion (Fig. 4). All intermediate and advanced users were able to complete the cardia biopsy task, while 1 of the 16 novices was unable to locate the lesion. All subjects were able to perform the simple fundal biopsy task.

### Questionnaire Data

The pre-simulator questionnaire data determined the subjects' previous experience level with computers and simulators. There were no significant differences between the groups in prior level of computer and simulator experience. The post-simulator questionnaire was designed to evaluate the experiences of the simulation subjects (Fig. 5). Subjects were asked to rate their opinions on a Likert scale from 1 to 5 (1 = strongly disapprove; 5 = strongly approve). Subjects were asked if they felt the simulator would be useful to them, at their current level of training, as well as whether the device would specifically be useful for teaching beginners. Novices and intermediates were the most satisfied with the VR experience; they felt that the simulation offered a valuable learning experience, especially for manipulation of the scope. Generally, all users felt the device would be useful for novices. Advanced users commented on the lack of realism of the simulator for their own use. They commented that the lack of realism in the graphical model, including the absence of enteric mucus, the ability to perform suction and insufflation, and the absence of obstructive



**Figure 5.** User attitudes towards the simulation experience assessed on a standard Likert scale (1–5; 1 = strongly disapprove, 5 = strongly approve).

landmarks such as the pharynx, upper esophageal sphincter, and pylorus made this simulation too easy. The novices and intermediates felt that they would use this simulator in their spare time if it were available.

## DISCUSSION

The increasing challenges of educating residents in the modern surgical residency program suggest that new methods are needed. A surgeon's skills are developed during procedures on a limited number of patients. The quality of training received is influenced by chance patient encounters and subjective methods of skills assessment.

The integration of skills laboratories into the traditional "see one, do one, teach one" methodology of education has already begun in many institutions. Even if not part of a formal skills training curriculum, most institutions have knot-tying boards and simple laparoscopic trainers available for their residents to use. More complex advanced skills trainers employing VR may be the next generation of tools for surgical training outside the OR. Its objective and quantitative methods of scoring performance, coupled with visually engaging displays, make this technology attractive for use in skill trainers.

Applications of real-time simulation for training involving computer modeling have already been incorporated into fields such as air and space flight, large military and commercial vehicle control, mechanical systems maintenance, and nuclear power plant operations<sup>16</sup>—all fields where procedures are hazardous and mistakes costly. These systems have been used because they provide training in a controlled, secure, and safe environment. Medical procedures should be no exception.

Before simulators are widely accepted and implemented into training programs, they must first be evaluated for their

validity as teaching devices. This study aimed to determine the construct validity of the 5DT Upper GI simulator as the first step in the validation of a device as a teaching tool. It asks the question: Can this training tool discriminate between users of differing skill levels performing routine tasks? In this study, groups with more endoscopic experience performed better than their less experienced counterparts in both the ability to visualize the upper GI tract as well as in performance of biopsy tasks. By using a subset of the internal metrics supplied by the system, we can demonstrate differences in the performance of groups of subjects with differing skill levels.

In the visualization task, the user's level of experience varied inversely with the time it took to inspect the upper GI tract: more advanced users finished the task quicker. Further, the amount of surface area viewed was directly related to the user's experience: more advanced users inspected more area. These results are not independent but rather multiplicative: advanced users visualized the GI tract more completely and did so in less time. If one were to establish an efficiency ratio, defined as area viewed/time, the differences between the groups would become even more pronounced. This finding is consistent with how we would expect a more experienced user to perform. More experienced users were better able to navigate past the pylorus and therefore covered more surface area. This is consistent with their having acquired the skill of paradoxically advancing the tip of the instrument through the pylorus and having to advance the tip cephalad while moving the scope caudally. Performance on the visualization tasks is suggestive of a measure of scope-maneuvering dexterity and technical ability to navigate the scope through the upper GI tract.

The user's performance on the biopsy task is a result of several factors: the ability to navigate, the ability to control the scope, coordination of the scope with the biopsy tool, overall dexterity, and to a degree, *a priori* knowledge of how to systematically look for lesions. This more complex task tests several skills in concert, so results are harder to attribute to any one single skill. As a successful procedure requires the coordination of many different skills, the test is a more complete test of users' overall abilities.

Performance differences between groups were also demonstrated in two of the three biopsy tasks. In biopsies 2 and 3, the biopsies of duodenum and cardia, advanced users took less time to find and biopsy the lesions. They also did so with fewer wall collisions. A direct relationship was seen for both criteria. As with the efficiency ratio described above, if one were to define an agility score as  $1/(\text{time to complete task} * \text{number of collisions})$ , the results also become more pronounced.

Biopsy 1 failed to demonstrate these trends. The fundal lesion of the first biopsy was directly in view on entering the stomach. In contrast, both the second and third lesions were harder to find. We believe that these lesions challenged the users' ability to perform endoscopy to a much greater level

than the fundal lesion, and therefore statistical differences were seen only on the latter two lesions.

The first lesion was designed to be accessible to everyone, to ensure that all subjects had the necessary baseline skills to access and biopsy lesions. That differences between the performance of the groups were not seen in this easier task suggests that a certain degree of difficulty is required to test user skills, and that this level was not achieved on the readily accessible “straight-shot” fundal biopsy.

Additionally, the other lesions were placed in locations that required some a priori knowledge of where to look. The duodenal lesion, suggestive of an ulcer, could not be readily seen without entering the pylorus. The cardia lesion, suggestive of a malignancy, needed to be sought through retroflexion back onto the cardia. Novice users might not know to maneuver the scope into the pylorus or to retroflex to look for the cardia lesion. More experienced users performed this maneuver as one step in their checklist of tasks during a routine endoscopy and had the procedural skills to access the lesions. The cognitive knowledge about where to look for common lesions and a knowledge-based framework for examining the enteric tract in an orderly fashion are attributes not directly tested by this simulation. Rather, they may be inferred as ability to complete the tasks quickly and thoroughly, if at all.

VR simulators provide a valid, reliable, and unbiased assessment of skills to identify specific weaknesses and strengths of users and to create individual profiles. Our study suggests a role for such a simulation device in identifying areas of difficulty for novices and providing a safe environment for them to practice manipulation skills. The steps of a complete examination, such as an upper endoscopy, can be taught on a computer simulation instead of on a live patient. Basic dial control of the gastroscope can be learned, especially through the more complicated anatomy, in a safe and time-independent fashion.

One interesting observation that can be made from this and other previously unpublished data by our group is that users of differing skills levels did not like the simulation equally and focused their comments in different areas. Both novices and intermediates praised the device as a useful teaching tool, especially as a means to teach scope manipulation. Advanced users, however, felt the system did not provide a realistic endoscopy experience. They commented that the model’s lack of a difficult oropharynx to instrument and the lack of an anatomically distinct upper esophageal sphincter and pylorus made the navigation tasks too easy. The lack of mucus to obfuscate the view, as well as lack of need to insufflate the esophagus, made the overall simulation too easy and in fact ignored what they felt to be the harder steps of the actual procedure. Clearly, users of differing experience levels had different expectations as to what the simulation should entail. Novices concentrated more intently on navigation tasks, while the more experienced users who had already mastered these skills were free to focus on issues of graphical and procedural realism and to

notice their absence. This higher level of detail may be important to experts’ intraprocedural decision-making abilities, skills that the novices would not have yet acquired.

These differences beg the questions: What is the best role of simulation devices? Can one simulation device be all things to all people? Do devices fit the need best as specific skill trainers, or as whole procedure simulators? All users felt that this device performed the former role well; it was in the latter role that differences in opinion were discovered. This device and future generations of devices will continue to improve. Feedback from expert users (usability) is important as successive iterations of these devices more accurately replicate realistic clinical scenarios. As the graphical images and haptic devices become more complex, their ability to simulate with greater detail will grow. Video games now available offer an astounding level of graphical complexity and employ force feedback to improve the user’s immersive experience, and similar programming techniques may be applicable to medical simulations. In the meantime, however, the ability for devices such as this to teach procedural techniques such as navigation and manipulation is already realized.

The ability to generate an objective report card of the user’s performance is a powerful learning tool. Many of the less experienced users were interested in reviewing their procedural mistakes and trying again with supervision after they had been run through the study. For remediation, beginners can immediately identify their weaknesses and practice those skills repeatedly in a safe manner. Identification of technical mistakes, omissions, and inaccuracies can be made. Improvements can be tracked over time as subjects practice on both the simulator and clinical patients. The objective analysis of performance afforded by computers may be useful to residents in training. A recent study demonstrated the inaccuracies intrinsic in self-assessment of procedural skills, as subjects tended to overcredit themselves.<sup>17</sup> Additionally, objective evaluations of procedural dexterity used in concert with subjective grading may offer the most useful and acceptable combination.<sup>18</sup>

Whether similar objective skills assessment measurements will become the basis of a grading system for users is yet to be seen. It is easy to envision a role for the integration of VR simulators into the educational curriculum of residents in training, as part of an extension of the skills laboratories in place in many institutions. In a further step, these skill assessment devices may be used to objectively assess and follow the performance of residents during training. On an even greater level, it is conceivable to imagine similar devices being used to test for a minimum level of skill proficiency in an objective manner as part of the credentialing process for specific procedures. Similar to this would be the use of VR devices as part of the certification and recertification of attending-level practitioners. More progress in simulation will have to be made to accurately model the complexities that would be tested. That progress, however, is sure to come.

Further work will include studies to determine the transferability of the skills learned in VR to users' performance on real patients. Once this is accomplished, real comparisons of cost savings of this method of teaching can be performed. The use of VR simulations may permit physicians to spend their time in the OR more safely, more effectively, and potentially more economically.

## References

1. Folse JR. Surgical education—addressing the challenges of change. *Surgery*. 1996;120:575–579.
2. Reznick RK. Teaching and testing technical skills. *Am J Surg*. 1993;165:358–361.
3. Bridges M, Diamond DL. The financial impact of teaching surgical residents in the operating room. *Am J Surg*. 1999;177:28–32.
4. Greenfield LG. Support of graduate medical education. *Curr Surg*. 1986;43:271.
5. Bridges M, Diamond D. The financial impact of teaching surgical residents in the operating room. *Am J Surg*. 1999;177:28–32.
6. O'Toole RV, Playter RR, Krummel TM, et al. Measuring and developing suturing technique with a virtual reality surgical simulator. *J Am Coll Surg*. 1999;189:114–117.
7. Taffinder N, Sutton C, Fishwick RJ, et al. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Studies in Health Technology and Informatics*. 1998;50:124–130.
8. Wanzel KR, Hamstra SJ, Anastakis DJ, et al. Effect of visual-spatial ability on learning of spatially-complex surgical skills. *Lancet*. 2002;359:230–231.
9. Watson DC, Matthews HR. Manual skills of trainee surgeons. *J R Coll Surg Edinb*. 1987;32:74–75.
10. Harris CJ, Herbert M, Steele RJ. Psychomotor skills of surgical trainees compared with those of different medical specialists. *Br J Surg*. 1994;81:382–383.
11. Squire D, Giachino AA, Proffitt AW, et al. Objective comparison of manual dexterity in physicians and surgeons. *Can J Surg*. 1989;32:467–470.
12. Steele RJ, Walder C, Herbert M. Psychomotor testing and the ability to perform an anastomosis in junior surgical trainees. *Br J Surg*. 1992;79:1065–1067.
13. DesCôteaux JG, Leclère H. Learning surgical technical skills. *Can J Surg*. 1995;38:33–38.
14. Gibbons RD, Baker RJ, Skinner DB. Field articulation testing: a predictor of technical skills in surgical residents. *J Surg Res*. 1986;41:53–57.
15. Satava RM. Surgical education and surgical simulation. *World J Surg*. 2001;25:1484–1489.
16. Gorman P, Lieser J, Murray W, et al. Evaluation and skill acquisition using a force feedback, virtual reality-based surgical trainer. *Medicine Meets Virtual Reality: Art, Science, and Technology*. Amsterdam, The Netherlands: IOS Press, 1999:121–123.
17. Evans AW, Aghabeigi B, Leeson R, et al. Are we really as good as we think we are? *Ann R Coll Surg Engl*. 2002;84:54–56.
18. Shah J, Darzi A. Surgical skills assessment: an ongoing debate. *BJU Int*. 2001;88:655–660.