

Development of a Model for Training and Evaluation of Laparoscopic Skills

Anna M. Derossis, MD, Gerald M. Fried, MD, Michal Abrahamowicz, PhD, Harvey H. Sigman, MD, Jeffrey S. Barkun, MD, Jonathan L. Meakins, MD, *Montreal, Quebec, Canada*

BACKGROUND: Interest in the training and evaluation of laparoscopic skills is extending beyond the realm of the operating room to the use of laparoscopic simulators. The purpose of this study was to develop a series of structured tasks to objectively measure laparoscopic skills. This model was then used to test for the effects of level of training and practice on performance.

METHODS: Forty-two subjects (6 each of surgical residents PGY1 to PGY5, 6 surgeons who practice laparoscopy and 6 who do not) were evaluated. Each subject viewed a 20-minute introductory video, then was tested performing 7 laparoscopic tasks (peg transfers, pattern cutting, clip and divide, endolooping, mesh placement and fixation, suturing with intracorporeal or extracorporeal knots). Performance was measured using a scoring system rewarding precision and speed. Each candidate repeated all 7 tasks and was rescored. Data were analyzed by linear regression to assess the relationship of performance with level of residency training for each task, and by ANOVA with repeated measures to test for effects of level of training, of repetition, and of the interaction between level of training and repetition on overall performance. Student's *t* test was used to evaluate differences between laparoscopic and nonlaparoscopic surgeons and between each of these groups and the PGY 5 level of surgical residents.

RESULTS: Significant predictors of overall performance were (a) level of training ($P = 0.002$), (b) repetition ($P < 0.0001$), and (c) interaction between level of training and practice ($P = 0.001$). There was also a significant interaction between level of training and the specific task on performance scores ($P = 0.006$). When each task was evaluated individually for the 30 residents, 4 of the 7 tasks (tasks 1, 2, 6, 7) showed significant correlation between PGY level and score. A sig-

nificant difference in performance scores between laparoscopic and nonlaparoscopic surgeons was seen for tasks 1, 2, and 6.

CONCLUSIONS: A model was developed to evaluate laparoscopic skills. Construct validity was demonstrated by measuring significant improvement in performance with increasing residency training, and with practice. Further validation will require correlation of performance in the model with skill in vivo. *Am J Surg.* 1998;175:482-487. © 1998 by Excerpta Medica, Inc.

Laparoscopic surgical training is multifaceted, and must include familiarity with laparoscopic instrumentation and mastering of fundamental skills necessary for the safe performance of laparoscopy. Differences between conventional and laparoscopic surgery include optics and instrumentation. Surgeons must learn to operate with long instruments, which amplify tremor and are harder to control than conventional instruments. These same instruments are limited in their range of motion by the trocars through which they must be passed. Furthermore, the constraints of length and width of these instruments have limited engineering design. Optical differences between open and laparoscopic surgery add to the difficulty in learning laparoscopy. Surgeons must learn to operate while looking at a monitor in another direction. Conventional laparoscopic imaging systems provide two-dimensional vision. Thus, depth perception is lacking, and surgeons must learn other cues to provide a sense of depth. These cues include the sense of touch, and the interaction of light and shading. More recently, first generation three-dimensional (3D) imaging systems have been introduced into laparoscopic surgery. Preliminary studies with small groups have not yet clearly suggested the benefits of such a system.^{1,2} There are some limitations of this generation of 3D laparoscopes. These include the need to wear glasses or goggles, and the problem with focusing due to the fact that the anthropometric interpupillary distance and accommodation have not yet been incorporated into the system.³

Videoendoscopic surgery requires ambidexterity, hand-eye coordination, and depth perception. Many methods have been used for training and credentialing. These include courses with practicums in animal models and training with other certified laparoscopic surgeons. One of the models utilized for training in laparoscopy is the surgical simulator, which allows the operator to enhance his/her motor skills in a safe and controlled environment.

The purpose of this study was to develop a method to quantitatively assess technical skills in laparoscopic surgery by measuring performance through a series of exercises

From the Divisions of General Surgery and Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada.

This work was supported by an educational grant from United States Surgical Corporation (Auto Suture Canada).

Requests for reprints should be addressed to G. M. Fried, MD, Division of General Surgery, Montreal General Hospital, 1650 Cedar Avenue Rm L 9-412, Montreal, Quebec, Canada H3G 1A4.

Manuscript submitted July 21, 1997 and accepted in revised form January 12, 1998.

performed within a surgical simulator. These tasks were then used to evaluate the effects of level of training and practice on performance.

MATERIALS AND METHODS

The McGill Laparoscopic Simulation study involved 42 examinees (volunteers) in total. There were 6 residents from each of the 5 years of the general surgery residency program, 6 practicing laparoscopic surgeons, and 6 attending surgeons who do not practice laparoscopy. It was required that each examinee complete seven exercises in turn and then repeat them. The simulator consists of a laparoscopic trainer box measuring $40 \times 30 \times 19.5$ cm (USSC Laptrainer, United States Surgical Corporation, Norwalk, Connecticut) covered by an opaque membrane. Two 12-mm trocars (USSC Surgiport, United States Surgical Corporation, Norwalk, Connecticut) were placed through the membrane at convenient working angles on either side of the 10-mm zero degree laparoscope (USSC Surgiview, United States Surgical Corporation, Norwalk, Connecticut). Four alligator clips within the simulator were used to suspend materials for certain exercises. The laparoscope and camera (Storz endoskope; telecam) were mounted on a stand at a fixed focal length. This enabled the examinee to work independently. The optical system consists of the laparoscope, camera, light source and video monitor (Sony Trinitron 19 inch). The video monitor was placed in line with the operator.

Seven standardized exercises were devised ranging from basic to more advanced laparoscopic skills. Some of the exercises aimed to develop laparoscopic coordination skills, others emphasized the use of certain laparoscopic instruments, and others focused on particular laparoscopic techniques.

Performance of each task was scored for both precision of performance and speed. For each exercise a timing score was calculated by subtracting the time to complete the exercise from a preset cutoff time: timing score = cutoff time (seconds) minus time to complete the exercise (seconds). This system rewards faster performance with higher scores. If the time to complete the exercise surpassed the preset cutoff time, a timing score of "0" (zero) was given, as no negative values were assigned. Precision of performance was also objectively scored by calculating a penalty score for each exercise (see description of exercises). Finally a total score combining the two above scores was calculated for each exercise performed by subtracting the penalty from the timing score: total score = timing score minus penalty score. Thus, the more accurately and quicker a task was completed, the higher the score.

A 20-minute introductory video demonstrating proper performance of all exercises was shown to each candidate prior to testing. The tasks are illustrated in **Figures 1 to 7**. The tasks were the following:

Pegboard patterns (Figure 1). Through the use of two pegboards and six pegs, the operator was required to lift each peg from one pegboard with the left hand, transfer it to the right hand, and place it on the other pegboard. This was then reversed. The aim was to test eye-hand coordination and ambidexterity. This was scored by time to completion with a cutoff time of 300 sec, and a penalty score was given by calculating the percentage of pegs that

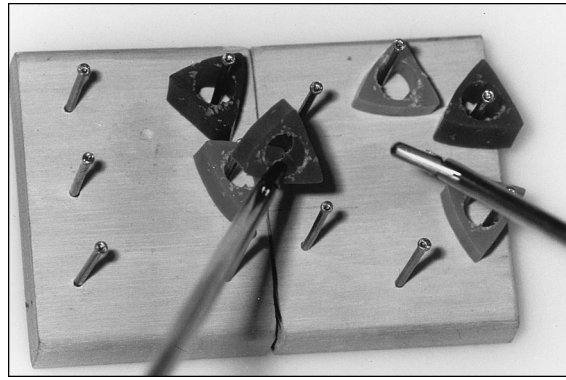


Figure 1. Pegboard patterns.

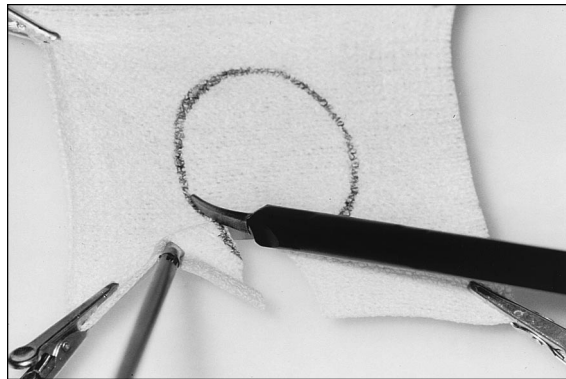


Figure 2. Pattern cutting.

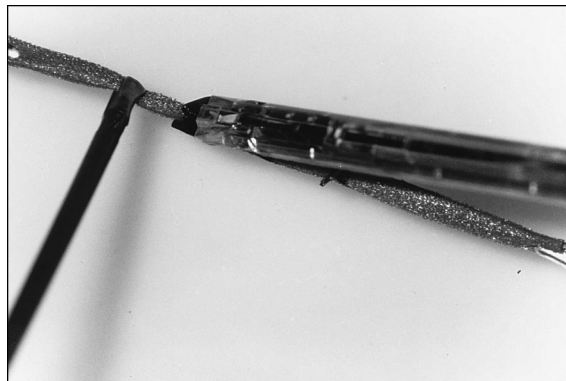


Figure 3. Clip application.

could not be transferred as a result of being dropped outside the field of view.

Pattern cutting (Figure 2). This task involved cutting a 4-cm diameter circular pattern out of a 10×10 cm piece of gauze suspended between alligator clips. The aim is to use the grasper in one hand placing the material under tension while cutting with the endoscopic scissors in the other hand. The score was determined by time to completion with a cutoff time of 300 sec, and a penalty was determined by calculating the percentage area of deviation from a perfect circle.

Clip application (Figure 3). This task involved placing two hemostatic clips on a tubular foam structure at pre-marked positions 3 cm apart, then cutting on a mark halfway between the clips. The purpose was to replicate

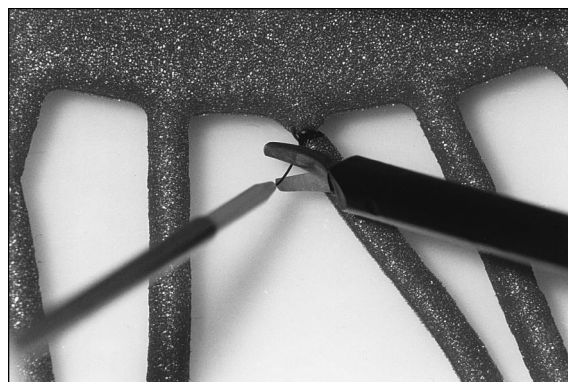


Figure 4. Placement of ligating loop.

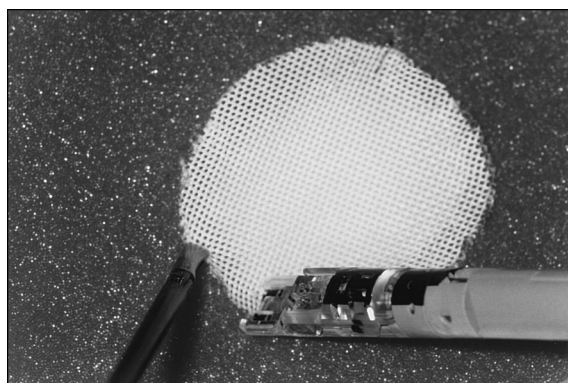


Figure 5. Mesh placement over a defect.

similar procedures in surgery where clips need to be placed accurately and securely. The score was determined by time to completion (cutoff time 120 sec), and the penalty score was calculated by measuring in millimeters the deviation of the clip or cut from the predrawn lines. A penalty score of 50 points was given for any clip not placed securely or completely across the tube.

Placement of ligating loop (Figure 4). This task involved the accurate placement and tightening of a commercially available pretied slip knot (USSC Surgitie, United States Surgical Corporation, Norwalk, Connecticut) on a foam tubular appendage. The aim was to replicate placement of a ligating loop such as might be required in an appendectomy. This involved backloading the ligating loop into a reducer, stabilizing the appendage, accurately and securely seating the knot, and cutting the excess suture. The score was determined by time to completion (cutoff time 180 sec) and the penalty score was calculated by measuring the distance in millimeters of the loop away from the premarked position. A 50-point penalty was given for any insecure or failed knot.

Mesh placement over a defect (Figure 5). This involved the placement of a 5-cm diameter mesh over a previously created 4-cm circular defect in a foam model, then securing the mesh to the foam with staples. The aim was to replicate mesh placement such as in hernia surgery. An allowance of four clips was used to secure the mesh. Scoring involved timing to completion (cutoff time 420 sec), and a penalty score was calculated. A penalty score of 25 points was given for each insecure clip, and 10 points for each extra

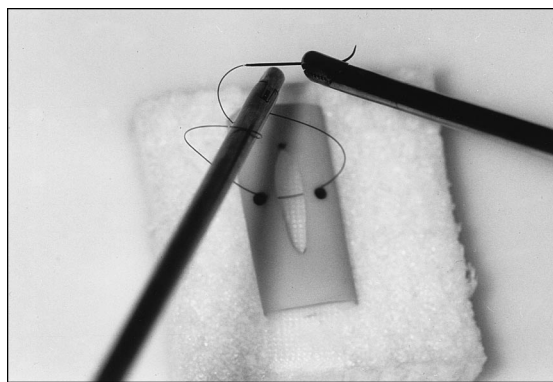


Figure 6. Intracorporeal knot.

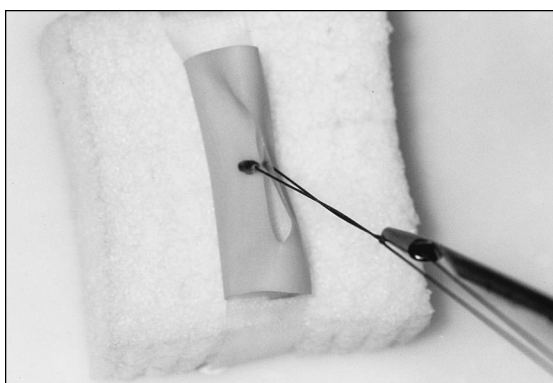


Figure 7. Extracorporeal knot.

clip used to a maximum of six. In the event where the mesh was misplaced leaving an area of defect, the percentage of defect was calculated and added to the penalty score.

Intracorporeal knot (Figure 6). This task involved the placement of a simple suture 13 cm in length through premarked points in a longitudinally slit Penrose drain. The suture was then tied using an intracorporeal knot technique. This exercise develops skills at needle transferring, placement of a suture, and knot tying. Scoring involved timing to completion (cutoff time 600 sec), and a penalty score reflecting the accuracy and security of the suture placed. The penalty score measured the distance in millimeters from the premarked points that the suture was placed; the gap in millimeters if the suture failed to approximate the slit; and the security of the knot, given 0 points for a secure knot, 10 points for a slipping knot, and 20 points for a knot that came apart. The sum of these three totaled the penalty score.

Extracorporeal knot (Figure 7). Identical to the previous task, except in this exercise the knot was tied outside the simulator and secured by an extracorporeal technique with a knot pusher. The scoring system used was identical to exercise 6, except the cutoff time was 420 sec.

Statistics

The overall data were analyzed with analysis of variance (ANOVA) to test for the effects of level of training and repetition (and their interaction) on performance. Linear regression analysis was used to relate total performance scores and timing scores for each task to residents' level of training. Correlation coefficients were calculated. Stu-

TABLE I

Postgraduate Year Versus Performance of Laparoscopic Task (Total Scores and Timing Scores)

		Task						
		1	2	3	4	5	6	7
Total score	<i>r</i>	0.62	0.69	0.30	0.34	0.26	0.42	0.45
	<i>P</i>	0.001	0.001	0.10	0.06	0.16	0.02	0.01
Timing score	<i>r</i>	0.62	0.69	0.22	0.34	0.21	0.42	0.44
	<i>P</i>	0.001	0.001	0.24	0.07	0.27	0.02	0.02

The correlation coefficients and *P* values of the total scores and timing scores for performance of each of the seven tasks versus level of training are summarized. Significant correlation was found for tasks 1, 2, 6, and 7 in the total score and timing scores.

TABLE II

Effects of Single Repetition on Performance (ANOVA)

	Task						
	1	2	3	4	5	6	7
<i>P</i> value	0.0001	0.0001	0.021	0.0001	0.005	0.0001	0.001

dent's *t*-test was used to test differences between laparoscopic and nonlaparoscopic attending surgeons, nonlaparoscopic surgeons versus PGY5 residents, and laparoscopic surgeons versus PGY5 residents. A value of $P < 0.05$ was considered significant.

RESULTS

The effect of level of training on performance was evaluated by ANOVA with repeated measures and was found to be a significant predictor of performance ($P = 0.002$). Linear regression analysis was used to correlate level of training with performance for each task individually. Overall four out of seven tasks (1, 2, 6, 7) showed a significant correlation ($P < 0.05$) between PGY level and total scores. Table I summarizes the correlation coefficients and *P* values for all seven exercises.

Timing score and penalty scores were also evaluated for each task. The improvement with postgraduate year was primarily attributed to the timing score (Table I). Correlation between level of training and timing scores was significant for tasks 1, 2, 6, and 7.

Linear regression analysis of the penalty scores versus postgraduate year was performed for tasks 2 and 5. These were found not to be significant. In the other tasks (1, 3, 4, 6, 7), errors were infrequent and did not lend themselves to statistical analysis (Table III).

There was a highly significant effect of a single repetition on overall performance by ANOVA ($P < 0.0001$). When each task was evaluated individually, all seven tasks showed a significant improvement in performance with repetition (Table II). The learning effect of repetition was found to be dependent on the level of training ($P = 0.001$). There was also a significant interaction between specific task and level of training on performance score ($P = 0.006$). A significant difference in total and timing scores between laparoscopic and nonlaparoscopic surgeons was seen for tasks 1, 2, and 6. Chief residents (PGY 5) performed better than nonlaparoscopic surgeons in tasks 1 and 6 (total score). In contrast, laparoscopic surgeons outperformed chief residents in tasks 2 and 3 (Table IV).

COMMENTS

Laparoscopic simulators have been used primarily as practicing tools. The simulator initially was constructed as a training model to assist surgeons in the development of coordination prior to application of a procedure in experimental animals.⁴ Although not specifically designed to simulate a specific surgical operation, the laparoscopic trainer provides fundamental training for most laparoscopic skills used in the majority of surgical operations. In addition, many surgeons may not have access to an animal laboratory facility. Practice should thus not be limited to intraoperative experience and a simulator in this setting is of great value.

A simulator allows for practice at various levels. Initially, one can start with looking directly through the transparent membrane using binocular vision. This will allow practice in hand coordination of instruments. One can then progress to working with the laparoscope and video monitor screen, with an opaque membrane.⁵ This adds the limitations of indirect vision. Sackier et al⁴ developed certain exercises that included suspending a bunch of grapes within the trainer and required the surgeon to identify a given grape and remove it without damaging the surrounding ones. Another exercise required the surgeon to carefully peel a tangerine using forceps and electrocautery. Others have described the use of long saphenous veins (removed during routine varicose vein surgery) as a model for training in laparoscopic exploration of the common bile duct.⁶ More recently, Rosser et al⁷ devised standardized drills and measured laparoscopic skill using timing alone as an endpoint. Trainees were evaluated in the performance of an intracorporeal suture before and after training with the drills. Progressive improvement in performing the drills was associated with steady acquisition of the suturing skill. In this study we developed a series of tasks ranging in difficulty and technique, and utilized the simulator to objectively evaluate laparoscopic performance at different levels of residency training and between laparoscopically trained surgeons and attending surgeons without laparoscopic expertise.

In evaluating the effect of level of training, there was a highly significant effect on overall performance. When each task was analyzed individually by linear regression relating performance to resident level, four out of seven tasks (1, 2, 6, 7) showed significant correlation with the total scores and timing scores. It is unclear why the other three tasks showed no significant correlation. This may be

TABLE III

Timing, Penalty, and Total Scores (Mean \pm SEM) for All Tasks for PGY1 to PGY5 Residents

	Task						
	1	2	3	4	5	6	7
R1							
Timing score	53 \pm 25	62 \pm 19	77 \pm 4	56 \pm 11	216 \pm 22	139 \pm 61	79 \pm 22
Penalty score	0 \pm 0	14 \pm 2	8 \pm 8	0 \pm 0	38 \pm 12	5 \pm 4	19 \pm 16
Total score	53 \pm 25	52 \pm 18	69 \pm 7	56 \pm 11	178 \pm 19	134 \pm 60	77 \pm 23
R2							
Timing score	132 \pm 17	111 \pm 26	84 \pm 3	87 \pm 11	283 \pm 28	269 \pm 57	139 \pm 31
Penalty score	0 \pm 0	12 \pm 3	1 \pm 0	1 \pm 1	28 \pm 12	6 \pm 4	5 \pm 2
Total score	132 \pm 17	103 \pm 25	83 \pm 4	86 \pm 12	255 \pm 28	267 \pm 56	134 \pm 31
R3							
Timing score	132 \pm 21	120 \pm 27	83 \pm 2	79 \pm 19	244 \pm 30	276 \pm 43	228 \pm 8
Penalty score	0 \pm 0	12 \pm 2	1 \pm 1	1 \pm 1	28 \pm 8	5 \pm 3	2 \pm 2
Total score	132 \pm 21	107 \pm 27	83 \pm 2	79 \pm 19	216 \pm 32	272 \pm 45	226 \pm 9
R4							
Timing score	138 \pm 20	160 \pm 19	79 \pm 6	91 \pm 9	241 \pm 44	193 \pm 69	153 \pm 25
Penalty score	0 \pm 0	13 \pm 2	1 \pm 1	1 \pm 0	34 \pm 14	36 \pm 20	3 \pm 2
Total score	138 \pm 20	147 \pm 20	79 \pm 6	90 \pm 9	207 \pm 43	191 \pm 69	151 \pm 24
R5							
Timing score	179 \pm 4	199 \pm 9	87 \pm 3	93 \pm 12	297 \pm 40	393 \pm 19	195 \pm 41
Penalty score	0 \pm 0	11 \pm 1	1 \pm 1	0 \pm 0	14 \pm 9	4 \pm 3	17 \pm 17
Total score	179 \pm 4	188 \pm 7	85 \pm 3	93 \pm 12	283 \pm 45	390 \pm 19	195 \pm 41

TABLE IV

Timing Score and Total Score (Mean \pm SEM) Comparison for Laparoscopic Surgeons, Nonlaparoscopic Surgeons, PGY5 Residents

	Task						
	1	2	3	4	5	6	7
Nonlaparoscopic surgeons							
Timing score	150 \pm 1*	178 \pm 2*	91 \pm 3	84 \pm 12	260 \pm 26	150 \pm 6*†	208 \pm 22
Total score	145 \pm 1*†	167 \pm 1*	91 \pm 3	83 \pm 12	232 \pm 34	150 \pm 6*†	205 \pm 22
Laparoscopic surgeons							
Timing score	192 \pm 11	221 \pm 7	95 \pm 3	106 \pm 6	216 \pm 53	323 \pm 69	206 \pm 26
Total score	192 \pm 11	212 \pm 6	95 \pm 3	105 \pm 5	191 \pm 58	322 \pm 69	205 \pm 26
PGY 5							
Timing score	179 \pm 4	199 \pm 9*	87 \pm 3*	93 \pm 12	297 \pm 40	393 \pm 19	195 \pm 41
Total score	179 \pm 4	188 \pm 7*	85 \pm 3*	93 \pm 12	283 \pm 45	390 \pm 19	195 \pm 41

The performance (total and timing scores) of laparoscopic, nonlaparoscopic, and chief residents (PGY 5) for each task is summarized.

* Performance superior by laparoscopic surgeons $P < 0.05$.

† Performance superior by PGY 5 residents $P < 0.05$.

explained by the exercise being poorly designed, too easy or too difficult to discriminate between residency level, or by a scoring system that did not combine precision and timing appropriately.

In analyzing task 3 (clipping and cutting), no significant correlation was seen, and the scores seem to cluster for all levels of training. It was postulated that perhaps this exercise was too simple and performance scores did not discriminate between junior and senior residents. Task 4 also showed no significant correlation. This may be because endoloop placement is not a skill frequently used clinically, and all residents, junior and senior, attempted this exercise relatively inexperienced. Despite this lack of correlation, it was thought to be a useful, instructive exercise. Task 5 (stapling the mesh) may be too difficult and needs some

revision. The defect in the foam model was close in size to the mesh, and the foam model was not rigidly fixed, which made stapling difficult.

The improvement with PGY year was primarily due to speed, rather than accuracy. Even though the penalty scores did not correlate with PGY training, they probably served as a deterrent to performing the tasks quickly without attention to accuracy. The effect of repetition on performance overall and for each task individually was highly significant, confirming the simulator model as a good practice tool. Further studies will address the learning curve in an effort to determine how many repetitions are required before performance scores plateau.

Comparing laparoscopic with nonlaparoscopic surgeons, there was a significant difference in the total scores and

timing scores for tasks 1, 2, and 6 (Table IV). With the exception of task 7, these are the same tasks found to show significant correlation of performance with level of training in residents. Perhaps no significant difference was seen in task 7 because a good part of extracorporeal knot tying is a skill familiar to nonlaparoscopic surgeons. Despite this difference between laparoscopic and nonlaparoscopic surgeons, nonlaparoscopic surgeons performed well, implying that some open surgical skills are transferable to the performance of laparoscopic surgery. This was also suggested by the observation that PGY5 residents performed superiorly to nonlaparoscopic surgeons in tasks 1 and 6 only. Performance of laparoscopic surgeons compared to PGY5 residents was only superior for tasks 2 and 3.

In assessing validity, one must determine whether the model measures what it claims to measure. For a test to have construct validity, it must be able to demonstrate a difference in score between those with skill and those without skill. This model demonstrates construct validity since there was a statistically significant improvement in performance with increasing level of training. This model needs to be further validated by demonstrating a correlation with performance of analogous tasks using an in vivo animal model and patients. These further validation studies are essential before using this tool to evaluate surgical skills. To date, this model should only be regarded as an innovative tool for technical skill evaluation. It also fulfills the characteristics of practicality and standardization, important criteria in the implementation of an educational tool. An informal questionnaire given to all examinees after the testing reflected a unanimous approval of the laparoscopic simulator as a good educational tool.

The primary function of the simulator to date is that of a practicing tool. It provides the benefit of practice and mastering laparoscopic techniques and use of laparoscopic instrumentation in a controlled environment. This may be done at the leisure or time scheduling of residents or attending staff. In addition, in the era of financial constraints where each added minute of operative time places a strain on the health care system, the effects of such a system on training will have obvious benefits. Techniques can be taught and practiced outside the realm of the operating room. This is not to imply the elimination of the animal model, but rather to complement its use, because

animals are expensive and not accessible to all. In the future, once validated, a model such as this may be used to evaluate new optical systems and instruments. This model would establish an objective assessment of residents performance in the laparoscopic domain. Presently, residency evaluations are primarily subjective, and assessment of surgical technique is usually based on a staff surgeon's recollection of a resident's performance rather than an objective assessment. A simulator can be used to set norms for different levels of training, and can be used as an evaluation tool in the assessment of laparoscopic skill in resident evaluations. Efforts are under way to develop and implement more quantitative approaches to measuring technical skills in surgery.⁸

In conclusion, despite its obvious importance in surgical teaching programs, the assessment and training of technical ability has lagged behind.⁹ If the way of the future leads to computerized simulators and virtual reality, simulation models will approximate more closely the realism of the operating room, and thus will provide more valid tools in the training and evaluation of surgical techniques.

REFERENCES

1. Jones DB, Brewer JD, Soper NJ. The influence of three-dimensional video systems on laparoscopic task performance. *Surg Laparosc Endosc.* 1996;6:191-197.
2. Crosthwaite G, Chung T, Dunkley P, et al. Comparison of direct vision and electronic two- and three-dimensional display systems on surgical task efficiency in endoscopic surgery. *Br J Surg.* 1995; 82:849-851.
3. Satava RM. 3D vision technology applied to advanced minimally invasive surgery systems. *Surg Endosc.* 1993;7:429-431.
4. Sackier JM, Berci G, Paz Partlow M. A new training device for laparoscopic cholecystectomy. *Surg Endosc.* 1991;5:158-159.
5. Mughal M. A cheap laparoscopic surgery trainer. *Ann R Coll Surg Engl.* 1992;74:256-257.
6. Windsor JA. Laparoscopic exploration of the common bile duct: a training model. *J R Coll Surg Edinb.* 1993;38:48-49.
7. Rosser JC, Rosser LE, Savalgi RS. Skill Acquisition and assessment for laparoscopic surgery. *Arch Surg.* 1997;132:200-204.
8. Winkel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. *Am J Surg.* 1994;167:423-427.
9. Dent TL. Training, credentialing, and granting of clinical privileges for laparoscopic general surgery. *Am J Surg.* 1991;161:399-403.