Surg Endosc (2002) 16: 130–137
DOI: 10.1007/s00464-001-8113-y

Surgical Endoscopy and Other Interventional Techniques

Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying

A. M. Pearson, A. G. Gallagher, J. C. Rosser, R. M. Satava

Department of Surgery, Yale University School of Medicine, 40 Temple Street, Suite 3-A, New Haven, CT 06510, USA

Received: 31 March 2001/Accepted in final form: 10 May 2001/Online publication: 12 November 2001

Abstract
Background: We evaluated the effectiveness of five training methods—four structured and one unstructured—for teaching intracorporeal knot tying.
Methods: Forty-three graduate students without prior laparoscopic experience were randomly assigned to one of five training groups, and their performance in 10 intracorporeal knot tying trials was evaluated, using time to complete a knot as the outcome measure.
Results: The average knot tying times for the four structured groups were significantly faster than the unstructured group ($p < 0.0001$). Among the four structured groups, the Minimally Invasive Surgical Trainer—Virtual Reality (MIST-VR) and the box trainer drills showed the most rapid improvements. The MIST-VR improved average suturing time from trial one to trial two ($p = 0.05$), the box trainer drills group improved from trial one to trial four ($p = 0.01$), and the other two groups showed slower improvements. Statistically significant correlations were observed between scores on MIST-VR tasks and average knot tying times ($r > 0.7$, $p < 0.05$).
Conclusion: Structured training can be useful for the development of laparoscopic skills. MIST-VR is a valuable part of this training, particularly in the objective evaluation of performance.

Key words: Objective assessment — Structured training — Laparoscopic training — Surgical skills — Virtual reality — Knot tying

Since its introduction in the late 1980s, minimally invasive surgery has become the standard procedure for a variety of surgeries, including cholecystectomy, hernia repair, appendectomy, Nissen fundoplication, and other advanced procedures. Laparoscopic cholecystectomy has been shown to result in less pain and shorter hospital stays than the open method, although overall complication rates are not significantly reduced [11]. This is due in part to the significantly higher rate of complications among the first patients treated by a surgeon, suggesting that surgeons must progress along a learning curve before laparoscopic procedures can be done proficiently [41]. This learning curve is partially a result of the unnatural environment in which minimally invasive surgery is practiced. Challenges facing the laparoscopic surgeon include the loss of important depth cues due to the use of a two-dimensional display monitor, loss of tactile or haptic feedback due to the nature of laparoscopic instruments, and the "fulcrum -effect" created by the insertion of the instruments through the abdominal wall, which causes the instrument tips to move in the opposite direction as the surgeon's hand [9, 14] Investigators have explicitly demonstrated these learning curves using simplified psychomotor tasks that mimic some of the skills required for laparoscopic surgery [8, 36].

The traditional Halstedian apprenticeship model of "see one, do one, teach one" is no longer adequate to train surgeons, since good laparoscopic skills cannot be developed by merely watching an expert [16, 17]. Laparoscopic proficiency is only realized after sufficient practice in the minimally invasive environment. To this end, a variety of approaches have been developed to teach laparoscopic skills outside of the operating room; these methods include practicing on animal models or artificial tissues, training boxes, and virtual reality simulators. Although live animal models are currently the most realistic training environment available outside the operating room, they present financial and ethical problems and are illegal in many countries [46]. Artificial tissues are useful for practicing basic surgical skills, but the scope of activities that can be learned from their use is limited [45]. Due to these limitations, surgical instructors have looked for other methods to teach minimally invasive surgery skills.

Developing laparoscopic skills in training boxes has been shown to be effective, and this approach is used in a number of training programs. Rosser has developed a course

Correspondence to: R. M. Satava

that first develops basic skills using three coordination drills and then teaches the advanced laparoscopic skill of intracorporeal suturing [36]. He demonstrated that residents could be taught to suture as well as experienced surgeons in a 2 1/2 day program and emphasized the importance of training outside of the OR operating room (OR) [35]. Fried et al. developed the MISTELS (McGill Inanimate System for Training and Evaluation of Laparoscopic Skills) program, which includes seven drills based on skills used in laparoscopic surgery and scores performance based on speed and accuracy. They showed that MISTELS scores correlate with surgical experience, thus demonstrating the construct validity of the program [9]. They went on to show that practice improved performance in four of seven tasks, scores improved over a 2-year period in residency training and there was a strong correlation between MISTELS scores and performance in an in vivo model, further validating the program [7, 8, 12]. Practicing in box trainers has been shown to be an effective way to develop basic laparoscopic skills, although no study to date has attempted to show that these skills can be transferred to the operating room, in part due to the difficulty of assessing surgical skill.

Although it has been estimated that judgment comprises 75% of a successful operation and technical skill 25%, this latter factor has traditionally been overlooked in the evaluation of surgical trainees [32]. Reznick et al. have developed and validated OSATS (Objective Structured Assessment of Technical Skill for Surgical Residents) to objectively assess surgical technical skill. This exam consists of six 15-min tasks that evaluate basic surgical skills, such as bowel anastomosis and T-tube insertion, and can be performed on either live animals or at benchtop stations with similar results [26]. Both task-specific checklists and global rating scales were shown to be reliable and valid in the evaluation of residents, since there was high interrater agreement and scores correlated with level of experience [33]. Training on the bench stations was also shown to transfer well to performing similar procedures on a cadaver, suggesting that the skills developed in the simulation lab may transfer to the OR [1]. Eubanks et al. have developed a similar checklist approach for the evaluation of laparoscopic cholecystectomy, which they demonstrated to be both reliable and valid [11]. Although both of these evaluation methods have been shown to be effective, they rely on subjective human ratings that are subject to inconsistency.

Other investigators have attempted to develop mechanical or computerized techniques to measure surgical skill more objectively. Motion analysis of surgical tools during surgery using electromagnetic trackers has been shown to be a valid measure of surgical skill that is more precise than simply measuring time to complete a task or subjectively evaluating performance [39]. Performance on an advanced endoscopic psychomotor tester (ADEPT) in the training box environment has been shown to correlate well with the subjective evaluation of operative skill and may also serve as an aptitude test of surgical ability [24]. Approaches such as these help to fill the need for objective evaluation of surgical skill, and in the near future virtual reality (VR) may provide an even more thorough and flexible mode of surgical evaluation.

The list of potential benefits of realistic VR surgical simulators is extensive: objective assessment of surgical skill, decreased risk to patients as surgeons progress along the learning curve in VR rather than in the OR, simulation of any type of case or complication imaginable, standardization of residency training regardless of the type of patients that present to each teaching hospital, experimentation with new procedures in a safe environment, and the ability to practice an operation on patient-specific anatomy prior to the real operation [16, 21, 37, 46]. To date, a variety of VR surgical simulators have been developed, ranging from a simple simulation of laparoscopic drills that can run on a PC [50] to an advanced hepatic surgery simulator that runs on parallel graphics workstations [25]. So far, they include simulations of neurosurgery [29, 44], ophthalmic surgery [38], laparoscopic surgery [2, 27], knee and shoulder arthroscopy [23, 40], endoscopic sinus surgery [3, 48], open abdominal surgery [4] anastomosis [30], bronchoscopy [5, 10], and intravenous catheter insertion [31, 47]. The realism of these simulators varies greatly depending on the power—and thus the cost—of the hardware on which they run. There is a constant balance between visual fidelity, real-time response, and computing power (cost) that developers must keep in mind as they design simulators. At this point, no simulator provides an experience that truly recreates surgery, but experts predict that surgical simulators will have reached a level of performance where they will be acceptable as tools for testing and certification by 2005–10 [16].

As virtual reality surgical simulators are developed, they must be shown to be both instructionally effective—that is, able to teach the real skills needed for surgery—and valid in their evaluation of surgical skills. Although many simulators have been developed, only a few have been assessed in terms of their effectiveness in training and evaluation. Most of these validation efforts have compared experienced surgeons to those with less or no surgical experience to demonstrate construct validity—showing that the simulator is measuring the skill it is designed to measure. The results have been mixed, with an anastomosis simulator [28], a knee arthroscopy simulator [22], an endoscopic sinus surgery simulator [48], and a laparoscopic skills simulator [6] demonstrating significant differences among users with different surgical experience, whereas other simulators have failed to discriminate between those with different levels of ability [31, 40]. Much more than construct validity needs to be evaluated to show that a surgical simulator is effective. It must also be shown to have predictive validity (the task predicts future performance), content validity (the task measures the content that is desired to be measured), concurrent validity (performance on the task mirrors performance in the real environment), and face validity (the simulation resembles the real task). In addition, a simulator must also be reliable (precise) and instructionally effective (the skills that are developed transfer to the real environment) [32].

So far, only a limited number of studies have looked beyond construct validity in their evaluation of surgical simulators. An IV catheterization simulator was shown to lack construct and concurrent validity because it could not differentiate between users of varying experience and the scores on the simulator did not correlate with ability to start an IV. It also lacked instructional effectiveness, since training on the simulator did not lead to improved IV placement [31]. The Minimally Invasive Surgical Trainer–Virtual Reality (MIST-VR) laparoscopic simulator has been shown to possess construct validity, because experienced surgeons perform better than novices. It is also instructionally effective for a basic task, since surgeons who trained on MIST-VR were shown to be better at a laparoscopic cutting task that those who had not [6, 13]. The issue that is not addressed by all of these studies is how well training on a simulator transfers to a complex task performed in surgery. To evaluate this, we looked at the effect of training with the Rosser drills and MIST-VR on learning the complex task of intracorporeal knot tying using a structured method of instruction. The performance of those in the structured groups was also compared to the performance of a group that learned knot tying in an unstructured fashion.

Materials and methods

Participants

Forty-three graduate level students (average age, 23.7 ± 1.8 years) who had no previous experience using laparoscopic instruments and were initially unable to tie an intracorporeal knot in under 600 secs were randomly assigned to one of five groups, each with a different protocol for teaching laparosocpic intracorporeal knot tying. The five groups were MIST-VR ($n = 10$; age, 24.0 ± 1.3), Box Trainer Drills ($n = 8$; age, 24.2 ± 1.2), Self-Practice ($n = 8$; age, 23.6 ± 1.2), No Practice ($n = 8$; age, 23.4 ± 3.3), and Unstructured ($n = 9$; age, = 22.9 ± 1.5).

Procedure

Subjects were randomly assigned to either one of the four structured training groups (MIST-VR, Box Trainer Drills, Self-Practice, No Practice) or the Unstructured group. The structured groups all went through the same didactic sessions based on Rosser's knot tying algorithm, the only difference between them being the type of psychomotor training they received prior to the timed knot tying trials [36]. The Unstructured group did not receive explicit training in the knot tying algorithm; functionally, they represented a group attempting to teach themselves how to tie an intracorporeal knot using only observation and basic knot tying principles. Those assigned to the four structured groups were given a 10-min demonstration on how to perform an intracorporeal knot by the instructor. Those in the Unstructured group watched a video of a knot being tied and received no verbal instruction. All subjects then attempted to place a single intracorporeal knot in a foam pad with no further guidance from the instructor. The subjects who successfully performed the baseline knot in < 600 sec were removed from the analysis.

The group assigned to MIST-VR (Virtual Presence Ltd., London, England) performed three virtual reality tasks. The system runs on an HIQ PC with a 400 MHz Pentium II processor and a three-dimensional graphics accelerator card (3D Blaster Riva TNT 2 Ultra; Creative Labs, Singapore). The input device consists of two standard laparoscopic grasper handles attached to rods that run through gimbals mounted on a steel frame (Immersion, San Jose, CA, USA). The simulated laparoscopic environment consists of an opaque box with gridlines and is displayed on a l7-in color monitor. All of the tasks consist of manipulating simple geometric figures with the instruments. No force feedback is provided to the instruments. Each subject performed 10 sets of each of the three tasks; all 10 sets are performed before moving on to the next task. To complete each task, three repetitions of the drill had to be performed with each hand.

In the first task, Acquire and Place (hard configuration), a sphere must be grasped with the instrument, placed into a cube, and then released while still within the cube. The second task, Transfer and Place (medium configuration), requires the subject to grasp the sphere with one instrument, pass it to the other instrument, and then release it within the cube. The final task, Traversal (medium configuration), consists of grasping the top of a cylinder with one instrument and then sequentially grasping lower segments of the cylinder with alternating instruments until the bottom of the cylinder is reached.

The system records time to complete the task, economy of motion (total path taken/optimal path length), and total number of errors. Errors are recorded for a variety of mistakes, such as collisions between instruments, dropping the ball outside the cube, or placing the instrument tip into the wrong segment of the cylinder. Feedback, given to the subjects between each set, consisted of scores for each of the above categories broken down into right and left instruments, as well as an overall score that was calculated by summing time, economy, and error scores. The subjects could monitor their progress over time because feedback was generated as graphs that showed both the most recent as well as previous scores. Lower scores reflect better performance. These three tasks were chosen because they most closely mirror the drills performed by the Box Trainer Drills group.

The Box Trainer Drills group performed 10 sets of each of the three drills designed by Rosser et al. [36]. These exercises were all performed in a laparoscopic trainer (Surgi Trainer; US Surgical Corporation, NorwalK, CT, USA) using a monitor (Trinitron model no. PVM-1943MD; Sony, Tokyo, Japan), a telescope three chips, 0°, 10 mm) (Stryker Endoscopy, Santa Clara, CA, USA), and a medical video camera (model no. 777; Stryker Endoscopy). Ten sets of each drill were completed before moving on to the next one. Each drill is timed by the instructor, who also maintains quality control.

The first drill, the Slam Dunk, requires the nondominant hand to transfer 10 black-eyed peas from the floor of the trainer through a hole in a cylinder by dropping them from a height of 1 cm above the aperture. The next drill, the Cobra Rope, requires the subject to pass a 60 × 1/8 in rope from one hand to the next. The rope has colored bands that are 1 in on length and 4 in apart, and subjects are only allowed to grasp the rope on these colored segments. The coiled rope is initially grasped with the nondominant hand; the next colored segment is then grasped with the dominant hand and passed to the nondominant hand. This maneuver is repeated until the entire length of rope has been traversed. The first two drills were performed using traditional endograspers (Auto Suture Endo Grasp 5 mm; US Surgical Corporation). The final drill, the Terrible Triangle, consists of engaging a curved needle mounted on a grasper (Rosser Signature Series Needle Holder; Asian Medical Technologies, Kalamazoo, MI, USA) into a metal loop at the apex of a wooden triangle. The axis of the camera is kept parallel to the plane of the loop, making it impossible to see both sides of the loop simultaneously and requiring the subject to locate the opening of the loop using other depth cues. With the needle in place, the triangle is transferred across the field to a circular landing pad, where the triangle must be placed and the needle removed from the loop. Five triangles must be transferred to complete each set.

Following the initial knot tying trial, the subjects assigned to the Self-Practice group practiced knot tying for 30 min without receiving any further instruction. This task was performed in the trainer on a foam pad, using the needle holder and grasper. The subjects assigned to the No Practice group did no further laparoscopic exercises following their initial attempt to tie a knot. Those assigned to the Unstructured group watched a video showing two more knots being placed and were also given written instructions on how to perform a basic instrument tie.

The subjects in the Box Trainer Drills, MIST-VR, and Self-Practice groups completed their assigned exercises with self-paced intervals between trials of ≤5; each set of exercises took ~30 min of on-task training time. Subjects were then shown a 20-min CD-ROM that proceeded through each step of the intracorporeal knot tying process (*The Art of Laparoscopic Suturing.*by J. C. Rosser, 1995, New York; (Radcliffe Medical Press). They were permitted to ask the instructor questions throughout the presentation. Thereafter, they each placed 10 intracorporeal knots on a foam pad in the box trainer using the needle holder and grasper [36]. To keep all subjects on the same cognitive level and prevent failure due to forgetting the knot tying algorithm, the instructor answered questions and gave instruction throughout these exercises whenever the subjects were unable to proceed on their own. Each knot was timed from grasping the suture with the assist device to the completion of the knot. The Unstructured group did not watch the CD-ROM and received no guidance from the instructor during their 10 knot tying trials. They had access to any written notes they had taken while watching the video and to the written instructions on how to perform an instrument tie.

Statistical analysis

The average knot tying times for each of the 10 timed trials as well as the average time for all 10 trials were compared using two-factor analysis of variance (ANOVA) for repeated measures. Contrast comparisons of dif-

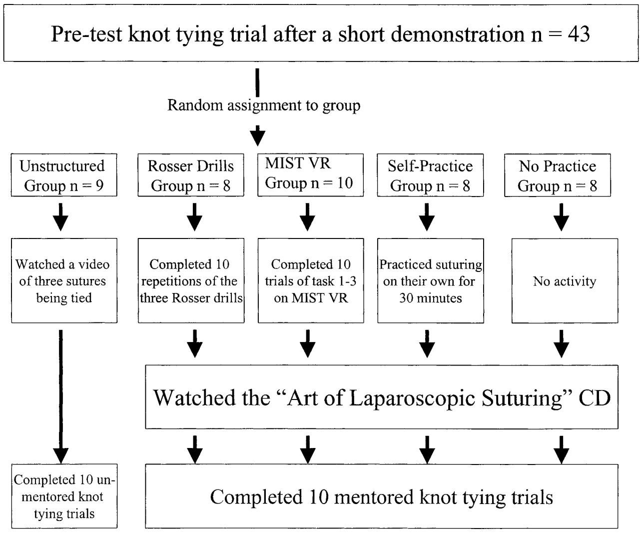

Fig. 1. Learning curve comparing baseline through trial 10.

ferences within each group between baseline and trial one, trial one and trial two, trial one and trial three, trial one and trial four, and trial one and trial 10 were made using one-factor ANOVA for repeated measures with the Scheffe F test. Correlations between scores on MIST-VR and times on the Rosser drills and average knot tying time were analyzed using the Pearson product-moment correlation coefficient. The threshold for significance for all statistical tests was set at $p = 0.05$ for a two-tailed test.

Results

The learning curves from the baseline trial through trial 10 for the five groups are shown in Fig. 1. Knot tying times averaged over the ten trials are shown in Table 1. There were statistically significant differences between the average knot tying times for each of the four structured groups (MIST-VR, Box Trainer Drills, Self-Practice, and No Practice) and the Unstructured group ($p < 0.0001$). Statistically significant differences between the structured groups and the Unstructured group were also seen for each of the 10 trials. There were no statistically significant differences among the four structured groups' average times over 10 trials or average times for any given trial. Table 1 also shows an analysis of differences between given trials within each group.

Significant decreases in time suggest that improvement has occurred in that interval. The greatest improvement was seen in the early trials, so differences between these trials were analyzed to determine the rate at which each group was improving. If there are decreases in time from one trial to the next, this indicates that improvement is rapid. On the other hand, if it takes multiple trials to produce a significant decrease in time, improvement must be slower. All four structured groups showed statistically significant improvement from the baseline trial to trial 1, whereas the Unstructured group did not. Only the MIST-VR group showed statistically significant improvement from trial 1 to trial 2 and from trial 1 to trial 3. The Box Trainer Drills group showed statistically significant improvement from trial 1 to trial 4. The Self-Practice group and the No Practice group showed no statistically significant improvement between trials 1 and 2, trials 1 and 3, or trials 1 and 4. All four structured groups showed significant improvement from trial 1 to trial 10. The Unstructured group showed no significant improvement throughout the knot tying trials.

Another approach to measure the effectiveness of a teaching method is to look at variability; decreasing variability within a group indicates that performance is becoming more consistent. A plot of standard error scores for the five groups over the course of the 10 trials is shown in Fig. 2. Although variable from one trial to the next, the four structured groups tended to show decreasing variability over the course of the 10 trials. In the first five knot tying trials, the Self-Practice group showed the lowest variability among the four structured groups, the No Practice group showed the greatest variability, and the MIST-VR and Box Trainer Drills groups showed intermediate levels of variability. The latter half of the trials showed relatively equal variability among the four structured groups. The Unstructured group's variability in the early trials is not comparable to the variability of the structured groups, since the vast majority of subjects in the Unstructured group failed to complete the early knot tying trials within the 10-min time limit. These subjects were given scores of 600 sec for these trials, and the high percentage of subjects receiving scores of 600 sec artificially decreased the variability in the early trials. As can be seen in Fig. 2, as some subjects learned how to successfully tie knots in the later trials, the variability in the Unstructured group increased to very high levels compared to the structured groups.

To look at how well scores on MIST-VR and times on the box trainer drills predicted knot tying ability, we calculated the Pearson product-moment correlation coefficient between these measures and knot tying times. Correlations between MIST-VR scores and knot tying times are shown in Table 2.

Statistically significant correlations were observed only with the Acquire and Place task. In this case, average error score, trial 1 time score, and trial 1 error score acted as strong predictors of knot tying performance. Scores on Transfer and Place and Traversal do not seem to predict knot tying performance. Weak correlations with knot tying times were observed for the Slam Dunk, the Terrible Triangle, and the Cobra Rope drills ($r = 0.03–0.66$), and these correlatios were not statistically significant.

Discussion

Structured training is an essential part of learning complex skills. The differences in performance between the structured groups and the Unstructured group make it clear that organized, step-by-step instruction is necessary to become proficient at the difficult task of intracorporeal knot tying. Though eight of the nine subjects in the Unstructured group were able to perform at least one successful intracorporeal knot in the 10 trials, their performance level was very poor in comparison to those who underwent structured training. To teach laparoscopic knot tying, one must develop a structured training method through task deconstruction and analysis, ultimately leading to the development of an algorithm that allows the task to be completed consistently and without difficulty. This study used Rosser's knot tying algorithm, which has been shown to be effective in teaching

Table 1. Analysis of knot tying learning curves

Group	Average Knot Time (sec)	Baseline vs trial 1		Trial 1 vs trial 2		Trial 1 vs trial 3		Trial 1 vs trial 4		Trial 1 vs trial 10	
		Scheffe F value	*p* value	Scheffe F value	*p* value	Scheffe F value	*p* value	Scheffe F value	*p* value	Scheffe F value	*p* value
MIST-VR ($n = 10$)	184.3 ± 71.5	3.71	0.001	2.08	0.05	3.55	0.001	2.68	0.01	6.43	0.0001
Box Trainer Drills ($n = 8$)	201.4 ± 51.7	2.11	0.05	1.42	n.s.	1.34	n.s.	2.61	0.01	4.47	0.0001
Self-Practice ($n = 8$)	153.6 ± 50.2	8.99	0.001	0.54	n.s.	1.15	n.s.	1.14	n.s.	2.64	0.01
No Practice ($n = 8$)	209.0 ± 77.7	2.65	0.05	0.31	n.s.	0.61	n.s	0.90	n.s.	3.29	0.01
Unstructured ($n = 9$)	448.4 ± 120.5	0.05	n.s.	0.03	n.s.	0.01	n.s.	0.17	n.s.	0.23	n.s.

n.s., not significant

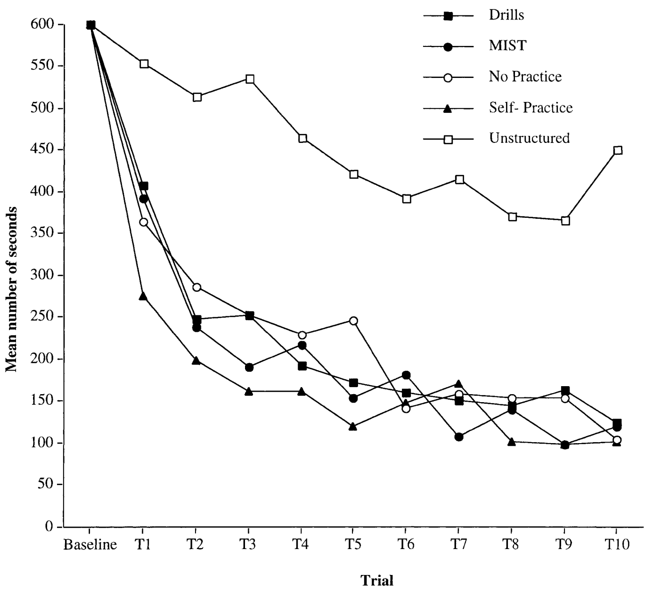

Fig. 2. The mean number of seconds taken by the five groups to tie an intracorporeal knot during the baseline and subsequent 10 trials.

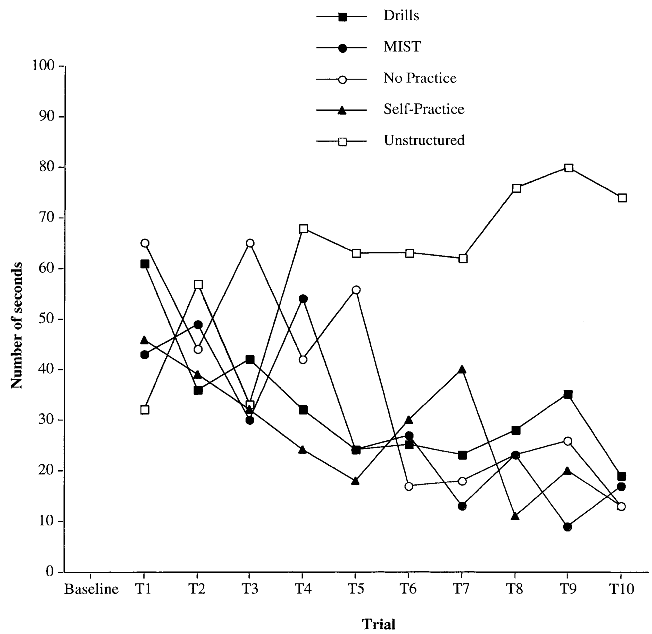

Fig. 3. The standard error scores for the five groups during the baseline trial and the subsequent 10 knot tying trials.

residents and experienced surgeons [35, 36]. The task was taught using both a CD-ROM shown to be effective in transferring cognitive knowledge and one-on-one instruction, with subjects reporting that both didactic approaches contributed to their learning [34]. The effectiveness of this algorithm has been validated, since students with no laparoscopic experience became proficient at intracorporeal knot tying in a few hours using this training method.

To our knowledge, the effects of performing coordination drills prior to learning a complex laparoscopic task such as intracorporeal knot tying have never been studied in a controlled experiment. There were no significant differences among the average knot tying times for the four structured groups, although the MIST-VR and Box Trainer Drills groups demonstrated more rapid improvement in the early knot tying trials. This finding suggests that the psychomotor skills developed in drills prior to the knot tying trials facilitated the early learning of the skill of laparoscopic knot tying, although they did not result in an increase in performance level at trial 10. Although 10 trials were used in this study for comparability with other studies, it appears that the learning curves had largely flattened out by trial 6.

Not surprisingly, all four structured groups showed significant improvement from the arbitrary baseline time of 600 secs to trial 1. This indicates that the training consisting of the CD-ROM and guidance from the instructor resulted in significant improvements in performance even for the No Practice group, which did not perform any coordination drills prior to knot tying. In looking at how improvement continues beyond trial 1, only the MIST-VR group and the Box Trainer Drills group showed statistically significant improvements from trial 1 to trial 4. This suggests that the psychomotor skills developed prior to knot tying led to more rapid improvement of knot tying skills. The MIST-VR group also showed statistically significant improvement from trial 1 to trial 2 and from trial 1 to trial 3, indicating that this group improved the most rapidly early in the trials. The Self-Practice group did not show statistically significant improvement from trial 1 to trial 4, suggesting that the early part of this group's learning curve was relatively flat. It may be that this group did not demonstrate greater improvement in the early trials because the subjects had already progressed down the learning curve during the practice session prior to the timed trials. This is reflected in the Self-Practice group tying the fastest average knots in the first five trials. The No Practice group failed to show significant improvement even from trial 1 to trial 4; thus, their rate of improvement in the early trials was slower than both

Table 2. Correlations between MIST-VR performance and average knot tying time

MIST-VR parameter	Correlation coefficient	t value	p value
Acquire and place task			
Average economy	0.056	0.16	n.s.
Average time	0.37	0.29	n.s.
Average errors	0.764	3.35	0.01
Trial 1 economy	0.54	1.81	n.s.
Trial 1 time	0.81	3.88	0.005
Trial 1 errors	0.729	3.01	0.017
Transfer and place task			
Average economy	0.37	1.13	n.s.
Average time	–0.091	–0.25	n.s.
Average errors	0.171	0.49	n.s.
Traversal task			
Average economy	0.319	0.95	n.s.
Average time	–0.12	–0.31	n.s.
Average errors	0.35	1.06	n.s.

n.s., not significant
All analyses were for nine degrees of freedom.

the MIST-VR group and the Box Trainer Drills group. By the end of the trials, the No Practice group performed at the same level as the other three structured groups. Thus, it appears that this group developed the psychomotor skills required for knot tying throughout the course of the 10 trials.

The No Practice group demonstrated the greatest variability among the structured groups in the first five trials, which suggests that training in the laparoscopic arena prior to learning knot tying was beneficial. The greater variability in this group indicates that in the early trials some subjects in this group were successfully learning knot tying while others were struggling. Performing psychomotor training prior to knot tying apparently led to more consistent learning throughout the MIST-VR, Box Trainer Drills, and Self-Practice groups. This may be a result of the training having substantially leveled the differences in psychomotor skills among the subjects prior to knot tying, which led to less variable knot tying times.

It is surprising that by the later trials the No Practice group reached the same level of performance as the groups that had received some type of training prior to knot tying. It appears that the CD-ROM, which teaches knot tying in a step-by-step manner, combined with one-on-one instruction in this technique can lead to success in learning how to tie knots, even without prior laparoscopic training. Knot tying is a complex task that requires cognitive learning of the knot tying algorithm as well as psychomotor skills. Even though the box trainer drills and MIST-VR can improve psychomotor skills, they obviously do not transfer the knowledge of the algorithm that is required to tie intracorporeal knots successfully [13, 19, 20]. This knowledge is clearly vital to learning how to tie knots, since the Unstructured group did not receive structured training in this method of knot tying and failed to learn how to tie them proficiently. It may be that cognitive learning is responsible for most of the improvement seen over the course of 10 trials, so that better psychomotor skills are not detectable simply by comparing average knot tying times. Although the psychomotor skills developed by performing the MIST-VR tasks and the box trainer drills led to greater improvement early in the learning process, the fact that the No Practice group was performing at the same level as the other groups by trial 10 suggests that these skills can be developed throughout the process as well.

Studies have shown that MIST-VR is an effective tool for developing the psychomotor skills required for laparoscopy [6, 13, 19, 20]. This study suggests that the skills developed on MIST-VR transfer to a complex laparoscopic task at least as well as the skills developed using the box trainer drills. MIST-VR also has the advantage of being able to be used by one individual without any assistance in holding the laparoscope or moving the objects used in the drills. Another very important feature of MIST-VR is its ability to objectively evaluate laparoscopic skills automatically, even in the absence of an instructor.

Previous studies have shown that MIST-VR is capable of discriminating between experienced surgeons, surgical residents, and individuals with no surgical experience [15, 43]. This study has shown that within a population of surgically naïve individuals, scores on particular MIST-VR parameters appear to be able to predict future laparoscopic success (as measured by ability to learn how to tie an intracorporeal knot). This result supports the findings of an earlier report showing that a subject's performance on MIST-VR could be predicted by the score obtained on the first trial [6]. These findings suggest that MIST-VR could be used both as an achievement test for practicing and training surgeons as well as an aptitude test for those considering surgery as a career. It appears that individuals possess varying levels of a number of skills that contribute to laparoscopic ability, and that MIST-VR is able to roughly determine their potential for laparoscopic success. Clearly, a large database of MIST-VR scores would have to be obtained before it could be useful as an aptitude test.

Now that a variety of training tools for laparoscopic surgery have been developed, their place in the surgical curriculum needs to be defined. The traditional surgical training model is a relatively unstructured one in which residents are expected to learn through observation and gradually develop skills as they get the opportunity to practice them. Our research, in concert with that of other investigators, suggests that such an unstructured approach is not effective for learning laparoscopy [16]. Instead, basic psychomotor skills should be developed first to provide a foundation for the acquisition of specific laparoscopic skills. Medical students and residents should initially develop basic skills using box trainer drill stations or virtual reality simulators, which have been shown to be effective in improving skills and are easy to use [35]. Once this foundation has been laid, surgical skills can be developed using task-specific drills such as MISTELS. These drills train for skills used in laparoscopy, such as clip application, ligating loop placement, mesh placement over a defect, and intracorporeal and extracorporeal suturing [9]. Box trainer versions of these drills are as effective as performing the drills in live animal models, so these drills also have the benefits of low cost and high accessibility [12]. After residents have become proficient in these skills essential to laparoscopy, they can proceed to learning different procedures in the animal lab. To maximize the effectiveness of this expensive training environment, residents must have already developed laparoscopic skills using inanimate models, so that the focus in the animal lab can be on learning specific procedures. For greatest effectiveness, these procedures should be taught in a structured manner prior to working in the animal lab. Once proficiency in the animal lab has been demonstrated, trainees should be ready to proceed to the OR with confidence in their ability to perform the procedure for which they have trained. For such a program to be effective, both faculty and trainees need to devote significant amounts of time and effort to the endeavor.

The importance of a structured curriculum with objective evaluation cannot be overemphasized. This study confirms the significance of the structured didactic lessons, such as the Rosser CD-ROM and live lectures. Throughout this program, objective feedback is essential so that both faculty and residents can evaluate performance and progress. Although both box trainer drills and MISTELS can be scored objectively, they cannot be measured precisely. The most objective evaluation tool currently available for measuring precision and economy of motion and accurately identifying errors is MIST-VR [13, 50].

Because it employs motion analysis, MIST-VR offers a much more precise and comprehensive evaluation of basic laparoscopic skills than can be measured by timing drills in box trainers. The benefits include an objective analysis of errors and economy of motion, two parameters that cannot be assessed accurately by an observer. Residents and faculty can monitor their progress on MIST-VR easily with real-time feedback and playback, and remedial training aimed at correcting specific deficiencies can be given to those who are not progressing well. It is possible that the number of residents who have difficulty becoming proficient at laparoscopy could also be reduced if evaluation tools like MIST-VR and ADEPT can be shown to be effective in screening out surgical candidates who lack the intrinsic ability to become good laparoscopic surgeons [24]. Objective assessment of performance in the animal lab and the OR is also essential. It can be accomplished through checklist evaluations for various procedures, as proposed by Eubanks et al. [11]. Therefore, we believe that a combination of structured training and objective evaluation would improve the effectiveness of laparoscopic training significantly.

Box trainers can be very effective in laparoscopic training, but the types of tasks they can simulate is limited. Even though basic skills can be taught in these trainers, more complex tasks, such as entire operations, cannot be simulated with a useful degree of realism. It is here that virtual reality trainers clearly have the potential to fill a void. Virtual organs and instruments would make it possible to perform entire "operations" without ever touching a patient. Since these tasks will be very specific—so specific that a given patient's anatomy can be modeled for practice—training in this environment should transfer very well to the operating room. Current limitations to realistic simulators include a shortage of funding to purchase high-end equipment and simulation software. Studies in the aerospace industry have shown that training on flight simulators significantly reduces the amount of flight time needed to learn a given skill and that total training time can actually be reduced by using simulators [18]. There is no reason to doubt that virtual reality surgical simulators will ultimately lead to similar results, so long as their development is encouraged. In the meantime, it is essential to continue the development of standardized objective measures of performance and a structured curriculum that these simulators can then incorporate.

Acknowledgments. This study was supported by a grant from Yale University School of Medicine, Medical Student Research Training Fellowship.

References

1. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, Hutchison C (1999) Assessment of technical skills transfer from the bench training model to the human model. Am J Surg 177: 167–170
2. Baur C, Guzzoni D, Georg O (1998) VIRGY: a virtual reality and force feedback based endoscopic surgery simulator. Stud Health Tech Inform 50: 110–116
3. Bockholt U, Ecke U, Muller W, Voss G (1999) Realtime simulation of tissue deformation for the nasal endoscopy simulator. Stud Heath Tech Inform 62: 74–75
4. Bro-Nielsen M, Helfrick D, Glass B, Zeng X, Connacher H (1998) VR simulation of abdominal trauma surgery. Stud Health Tech Inform 50: 117–123
5. Bro-Nielsen M, Tasto JL, Cunningham R, Merril GL (1999) Preop endoscopic simulator: a pc-based immersive training system for bronchoscopy. Stud Health Tech Inform 62: 76–82
6. Chaudry 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. Ann R Coll Surg Engl 81: 281–286
7. Derossis AM, Antoniuk MA, Fried GM (1999) Evaluation of laparoscopic skills: a 2-yr follow up during residency training. Can J Surg 42: 293–296
8. Derossis AM, Bothwell J, Sigman HH, Fried GM (1998) The effect of practice on performance in a laparoscopic simulator. Surg Endosc 12: 1117–1120
9. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL (1998) Development of a model for training and evaluation of laparoscopic skills. Am J Surg 175: 452–457
10. Englmeier KH, Haubner M, Krapichler C, Reiser M (1999) A new hybrid renderer for virtual bronchoscopy. Stud Health Tech Inform 62: 109–115
11. Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S, Pellegrini C (1999) An objective scoring system for laparoscopic cholecystectomy. J Am Coll Surg 189: 566–574
12. Fried GM, Derossis AM, Bothwell J, Sigman HH (1999) Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. Surg Endosc 13: 1077–1081
13. Gallagher AG, McClure N, McGuigan J, Crothers I, Browning I (1999) Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). Endoscopy 31: 310–313
14. Gallagher AG, McClure N, McGuigan J, Ritchie K, Sheehy NP (1998) An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. Endoscopy 30: 617–620
15. Gallagher AG, Ritchie K, McClure N, McGuigan J ( ) Objective psychomotor skills assessment of experienced, junior and novice laparoscopists with virtual reality. World J Surg : in press
16. Haluck RS, Krummel TM (2000) Computers and virtual reality for surgical education in the 21st century. Arch Surg 135: 786–791
17. Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, Waugh RA, Brown DD, Safford RR, Gessner IH (1999) Simulation technology for healthcare professional skills training and assessment. JAMA 282: 861–866
18. Jones MB, Kennedy RS (1996) Isoperformance curves in applied psychology. Human Factors 38: 167–182
19. Jordan JA, Gallagher AG, McClure N, McGuigan J ( ) Virtual reality training leads to faster automation of the novel psychomotor restrictions encountered by laparoscopic surgeons. Surg Endosc : in press
20. Jordan JA, Gallagher AG, McGlade K, McGuigan J (2000) A comparison between randomly alternating imaging, normal laparoscopic

imaging, and virtual reality training in laparoscopic skill acquisition. Am J Surg 180: 208–211
21. Kuppersmith RB, Johnston R, Jones SB, Jenkins HA (1996) Virtual reality surgical simulation and otolaryngology. Arch Otolaryngol Head Neck Surg 122: 1297–1298
22. McCarthy A, Harley P, Smallwood R (1999) Virtual arthroscopy training: do the virtual skills developed match the real skills required? Stud Health Tech Inform 62: 221–227
23. McCarthy AD, Hollands RJ (1998) A commercially viable virtual reality knee arthroscopy training system. Stud Health Tech Inform 50: 302–308
24. Macmillan AIM, Cuschieri A (1999) Assessment of innate ability and skills for endoscopic manipulations by the advanced Dundee endoscopic psychomotor tester: predictive and concurrent validity. Am J Surg 177: 274–277
25. Marescaux J, Clement JM, Tassetti V, Koehl C, Cotin S, Russier Y, Mutter D, Delingette H, Ayache N (1998) Virtual reality applied to hepatic surgery simulation: the next revolution. Ann Surg 228: 627–634
26. Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 84: 273–278
27. Noar MD (1995) The next generation of endoscopy simulation: minimally invasive surgical skills simulation. Endoscopy 27: 81–85
28. O'Toole RV, Playter RR, Krummel TM, Blank WC, Cornelius NH, Robertsn WR, Bell WJ, Raibert M (1999) Measuring and developing suturing technique with a virtual reality surgical simulator. J Am Coll Surg 189: 114–128
29. Phillips NI, John NW (2000) Web-based surgical simulation for ventricular catheterization. Neurosurg 46: 933–936
30. Playter R, Raibert M (1997) A virtual surgery simulator using advanced haptic feedback. Min Invas Ther Appl Technol 6: 117–121
31. Prystowsky JB, Regehr G, Rogers DA, Loan JP, Hiemenz LL, Smith KM (1999) A virtual reality module for intravenous catheter placement. Am J Surg 177: 171–175
32. Reznick RK (1993) Teaching and testing technical skills. Am J Surg 165: 358–361
33. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W (1996) Testing technical skill via an innovative 'bench station' examination. Am J Surg 172: 226–230
34. Rosser JC, Herman B, Risucci DA, Murayama M, Rosser LE, Merrell RC (2000) Effectiveness of a CD-ROM multimedia tutorial in transferring cognitive knowledge essential for laparoscopic skill training. Am J Surg 179: 320–324
35. Rosser JC, Rosser LE, Savalgi RS (1998) Objective evaluation of a laparoscopic surgical skill program for residents and senior surgeons. Arch Surg 133: 657–661
36. Rosser JC, Rosser LE, Savalgi RS (1997) Skill acquisition and assessment for laparoscopic surgery. Arch Surg 132: 200–204
37. Satava RM (1996) Advanced simulation technologies for surgical education. Bull Am Coll Surg 81: 77–81
38. Sinclair MJ, Peifer JW, Haleblian R, Luxenberg MN, Green K, Hull DS (1995) Computer simulated eye surgery. Opthalmology 102: 517–521
39. Smith S, Torkington J, Darzi A (1999) Objective assessment of surgical dexterity using simulators. Hosp Med. 60: 672–675
40. Smith S, Wan A, Taffinder N, Read S, Emergy R, Darzi A (1999) Early experience and validation work with procedicus VA—the prosolvia virtual reality shoulder arthroscopy trainer. Stud Health Inform Tech 62: 337–343
41. Southern Surgeons Club A prospective analysis of 1518 laparoscopic cholecystectomies. New Engl J Med 324: 1073–1078
42. Spencer F (1979) Competence and compassion: two qualities of surgical excellence. Bull Am Coll Surg 64: 15–22
43. Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A (1993) A Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic trainer. Stud Health Tech Inform 50: 124–130
44. Tanaka H, Nakamura H, Tamaki E, Nariai T, Hirakawa K (1998) Brain surgery simulation system using vr technique and improvement of presence. Stud Health Tech Inform 50: 150–154
45. Thomas WEG, Lee PWR, Sutherland GT, Day RP (1996) A preliminary evaluation of an innovative synthetic soft tissue simulation module for use in basic surgical skills workshops. Ann R Coll Surg Eng (Suppl) 78: 268–271
46. Torkington J, Smith SGT, Rees BI, Darzi A (2000) The role of simulation in surgical training. Ann R Coll Surg Engl 82: 88–94
47. Ursino M, Tatso JL, Nguyen BH, Cunningham R, Merril GL (1999) Cathsim: an intravascular catheterization simulation on a pc. Stud Health Tech Inform 62: 360–366
48. Weghorst S, Airola C, Oppenheimer P (1998) Validation of the Madigan ESS simulator. Stud Health Tech Inform 62: 399–405
49. Wherry DC, Rob CG, Marohn MR, Rich NM (1994) An external audit of laparoscopic cholecystectomy performed in medical treatment facilities of the department of defense. Ann Surg 220: 626–634
50. Wilson MS, Middlebrook A, Sutton C, Stone R, McCloy RF (1997) MIST VR: a virtual reality trainer for laparoscopic surgery assesses performance. Ann R Coll Surg Engl 79: 403–404