



## Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using criteria levels based on expert performance

A. G. Gallagher,<sup>1</sup> A. B. Lederman,<sup>1</sup> K. McGlade,<sup>2</sup> R. M. Satava,<sup>3</sup> C. D. Smith<sup>1</sup>

<sup>1</sup> Emory Endosurgery Unit, Emory University School of Medicine, 1364 Clifton Road, NE, Atlanta, GA 30322, USA

<sup>2</sup> Department of General Practice, Queen's University, Dunluce Health Centre, 1 Dunluce Avenue, Belfast BT9 7HR, Ireland

<sup>3</sup> Department of Surgery, University of Washington Medical Center, BB430, Box 356410, 1959 NE Pacific Street, Seattle, WA 98195-6410, USA

Received: 1 April 2003/Accepted: 17 September 2003/Online publication: 19 March 2004

### Abstract

**Background:** Increasing constraints on the time and resources needed to train surgeons have led to a new emphasis on finding innovative ways to teach surgical skills outside the operating room. Virtual reality training has been proposed as a method to both instruct surgical students and evaluate the psychomotor components of minimally invasive surgery *ex vivo*.

**Methods:** The performance of 100 laparoscopic novices was compared to that of 12 experienced (> 50 minimally invasive procedures) and 12 inexperienced (< 10 minimally invasive procedures) laparoscopic surgeons. The values of the experienced surgeons' performance were used as benchmark comparators (or criterion measures). Each subject completed six tasks on the Minimally Invasive Surgical Trainer—Virtual Reality (MIST-VR) three times. The outcome measures were time to complete the task, number of errors, economy of instrument movement, and economy of diathermy.

**Results:** After three trials, the mean performance of the medical students approached that of the experienced surgeons. However, 7–27% of the scores of the students fell more than two SD below the mean scores of the experienced surgeons (the criterion level).

**Conclusions:** The MIST-VR system is capable of evaluating the psychomotor skills necessary in laparoscopic surgery and discriminating between experts and novices. Furthermore, although some novices improved their skills quickly, a subset had difficulty acquiring the psychomotor skills. The MIST-VR may be useful in identifying that subset of novices.

**Key words:** Minimal access surgery — Task performance — Psychomotor skills — Surgical simulator — Minimally Invasive Surgical Trainer—Virtual Reality (MIST-VR)

It is clear that the paradigm for the training and assessment of surgeons is changing. The introduction of new technologies, such as laparoscopy and robotics, over the past decade has created a need for new and different psychomotor and perceptual skills that must be mastered by both the practicing surgeon and the surgeon-in-training. The performance of laparoscopy requires the practitioner to overcome several well-known perceptual and psychomotor problems. To reach proficiency, the surgeon must meet the challenges of degraded tactile feedback, conflicting depth perception cues, and the fulcrum effect of the patient's abdominal wall, which inverts the mechanics of moving instruments (e.g., contraintuitively, leftward hand movement will move the tip of the instrument to the right) [3, 6, 7, 10–12].

Coupled with the growing demand for new skills, the pressures exerted by regulations limiting resident work hours, an increase in patient throughput, and economic stringencies have intensified the challenge of designing relevant curricula for resident training and continuing medical education. Many surgical residents now graduate with only the most basic laparoscopic skills and are unable to perform procedures more difficult than a laparoscopic cholecystectomy or appendectomy [9]. In many cases, attending surgeons have had no formal training in laparoscopy, or their training has been limited to short 2–3-day courses [4, 12]. Surveys have found that only 18% of residents believed that their residency was preparing them to perform advanced laparoscopic procedures and that only 25% of members of the faculty

were competent to teach these skills [2]. This limitation in laparoscopic ability is reflected in resident case logs. In 2001–2002, graduating chief residents performed an average of 136.3 laparoscopic procedures as the operative surgeon, 76% of which were either appendectomies or cholecystectomies. The average number of laparoscopic antireflux procedures, reported by chief residents was  $5.9 \pm 7$ , with the most common number reported being a mode of zero. The statistics for minimally invasive herniorrhaphy were similar, with an average of  $9.8 \pm 10$  cases but a mode of zero [18]. Clearly, despite receiving their primary surgical training in an age of advanced laparoscopic surgery, a majority of new surgeons are not instructed in these skills.

Advances in simulator technology have been forecast and widely touted since the advent of the modern age of laparoscopy [14, 15]; however, simulator technology is still in its infancy. In the meantime, a simple, visually abstract simulator, the Minimally Invasive Surgical Trainer—Virtual Reality (MIST-VR; Mentice Inc., 12463 Rancho Bernardo Road, #198, San Diego, CA, USA), has become available to fill the gap. Although it lacks haptic feedback, MIST-VR has a good-quality interface that is an acceptable alternative to box trainers [1]. MIST-VR has been validated as an effective tool for instructing trainees in the psychomotor components of laparoscopic surgery [12]. There is substantial evidence that subjects who receive training on MIST-VR significantly outperform case-matched control groups, as assessed by a validated tool [7, 11, 19]. The first evidence that virtual reality skills can translate into improved outcome in the operating room was recently published in a prospective, randomized, blinded trial. In this trial, surgical residents who were trained according to defined performance criteria on the MIST-VR were able to perform laparoscopic gallbladder dissection faster, with lower completion rates and fewer errors, when compared to a group who received standard surgical training [16].

The ability to assess technical performance is a critical aspect of training. Virtual reality training systems can easily collect data via their computer-based platforms, including information on instrument manipulation and error. Time to task completion alone is an inadequate measure of performance, yet in most of the laparoscopic training programs that have been described it is the only benchmark used [17]. The MIST-VR is able to assess not only time to task completion but also task-specific parameters, such as economy of movement for each hand, number of errors, and economy of diathermy. These metrics provide an objective assessment of psychomotor skills, with reproducible, robust results. Several studies have shown construct validity of the MIST-VR by demonstrating its ability to consistently distinguish between experienced and inexperienced surgeons [5, 8, 13].

The next step in the validation of the MIST-VR as an assessment tool is to prove its ability to distinguish between the performances of individuals with the same level of experience but with different psychomotor abilities. The ability to discriminate between trainees in terms of psychomotor ability, would enable training to

be tailored to individual needs and to a definable criterion level.

Another critical factor for training and assessment is to determine how long or to what performance level a student should train. It has been suggested that a benchmark (the “criterion level”) be defined as the mean performance of experienced surgeons [16].

## Methods

### *Subjects*

A total of 100 medical students with no laparoscopic operative experience participated in the study (mean age, 22 years; range, 19–28; male, 68; female, 32). Students were either in their 1st–4th years as medical students at Queen’s University (Belfast, Ireland) or in their 3rd year of medical school at Yale University (New Haven, CT, USA). All subjects had expressed an interest in a surgical career. None had any surgical experience beyond the level of a 3rd-year American medical student (observation of surgical procedures or retraction as guided by the operating surgeon). The data for these subjects were compared to data previously published by our group on 12 experienced laparoscopic surgeons who had performed >50 laparoscopic operations (mean age, 39 years; range, 30–52), 12 less experienced surgeons who had performed more than one but <10 laparoscopic procedures, and a control group of 12 university students (novices) who had no medical background [5]. Any therapeutic laparoscopic procedure was counted; the most common procedure was laparoscopic cholecystectomy.

### *Apparatus*

The MIST-VR system used in this study was based on a 200-MHz Pentium personal computer (PC) running Windows 95 with 32-Mb RAM and a Matrox Mystique 4-MB video card (Matrox Graphics Inc., 1055 St Regis Blvd., Dorval, Quebec, Canada). The laparoscopic interface was a standard Virtual Laparoscopic Interface frame set (Immersion Corporation, 801 Fox Lane, San Jose, CA, USA) unit, with the addition of a foot pedal for the diathermy tasks. This setup contained two laparoscopic instruments held in position-sensing gimballs with five degrees of freedom. The trials ran MIST-VR v. 1.2, which utilized the WorldToolKit version and Microsoft Direct 3D v. 3 graphics libraries. Frame rates averaged ~15 frames per second (fps), enabling movements to be translated from the real instruments to the virtual world in real time and viewed on a 17-in color monitor. The interior of a three-dimensional (3D) cube on the computer screen represents an accurately scaled operating volume of  $10\text{ cm}^3$ . The image zoom and size of the target objects can be varied. Target images appear within the operating volume according to the skill task and can be virtually grasped, manipulated, or cauterized. Each of the different tasks is recorded exactly as it is performed, enabling accurate and reliable assessment. The monitor was placed at eye level with the laparoscopic instruments and at standard surgical height between the subject and the monitor.

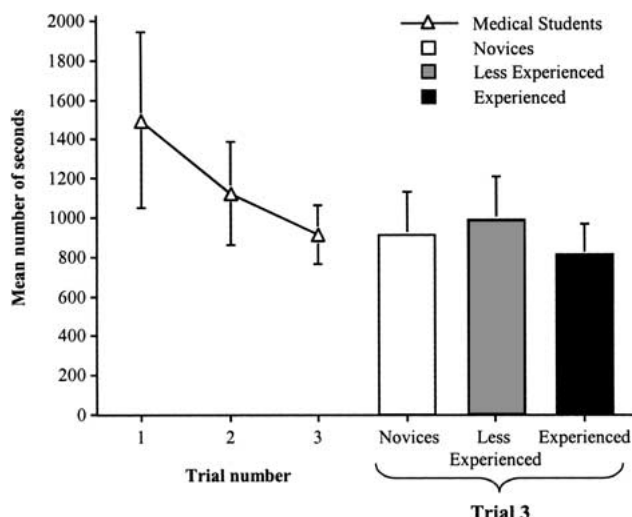
### *Procedure*

All subjects received supervised MIST-VR testing and completed all six simulation tasks (Table 1). Testing was completed in a quiet room near the operating rooms. There were five measures of the participants’ performance: time to complete all six MIST-VR tasks, number of errors, economy of movement of the right instrument, economy of movement of the left instrument, and economy of diathermy use during tasks five and six. All tasks were completed five times per trial for each hand.

*Time* was measured from the start of the task to the completion of the last sequence of movements of that task. The clock stopped automatically at the end of the task. The time that elapsed between the different tasks was not included.

**Table 1.** Tasks of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) and real-world correlation

Task	Description	Operative corollary
Acquire–place	Grab a sphere and place it in a wire-frame cube.	Grabbing tissue and retracting it to a given position
Transfer–place	Grab a sphere with one hand, pass it to the other hand, and place it in a wire-frame cube.	Manipulation of a needle during intracorporeal suturing
Traversal	Use instruments to travel from top to bottom along the outside of a three-dimensional cylinder.	Running the small intestine
Withdraw–insert	Pull an instrument out of the operative field and reinsert it for further use.	Withdrawing and placing new instruments
Diathermy	Cauterize a series of targets located on a sphere fixed in space.	Cauterizing a bleeding vessel
Manipulation–Diathermy	Grab a sphere and touch it with the instrument in the opposite hand; withdraw and reinsert the opposing instrument; apply diathermy to targets on the sphere while holding the sphere steady in a wire-frame cube.	Cauterizing the gallbladder off a liver bed during a laparoscopic cholecystectomy

**Fig. 1.** Mean number of seconds ( $\pm$ SD) taken by the medical students to complete the tasks in trials one through three in comparison to the control, less experienced, and experienced groups on trial three.

*Economy of movement* was assessed for each hand separately. The computer can easily measure the optimal distance the instrument needed to travel from its start to the target, as well as the actual distance traveled, and therefore any excess distance traveled. Economy of movement was defined as the ratio of the excess distance traveled by the instrument tip to the optimal distance.

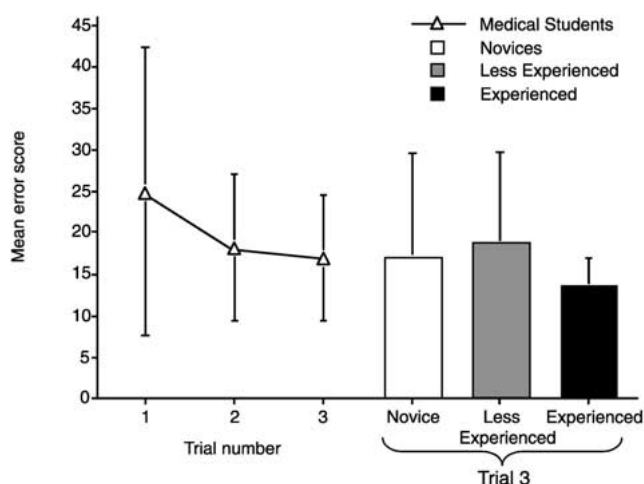
*Average error* was measured as the number of errors per task segment. For each task, the software applied an SE matrix.

The *economy of diathermy* was defined by the excess burn time divided by the optimal burn time. Tasks five and six were the only ones to use diathermy.

## Results

The results from the three trials completed by the medical students showed significant improvement in each metric with each trial, as well as a reduction in variability. There was also a reduction in the variability of their performance on the MIST-VR with each trial, seen as a decrease in SD with each trial. Overall, as a group, the medical students quickly climbed the learning curve.

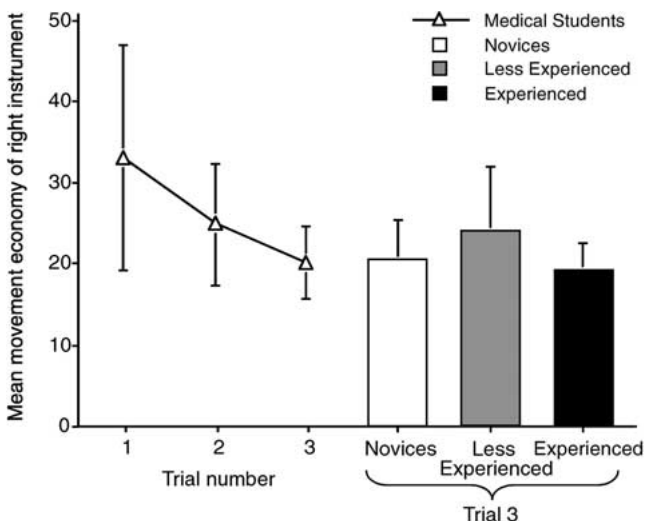
The medical students showed improved times with each trial, but at trial three they were still significantly

**Fig. 2.** Mean error scores ( $\pm$ SD) for the medical students in trials one through three in comparison to novice, less experienced, and experienced groups in trial three.

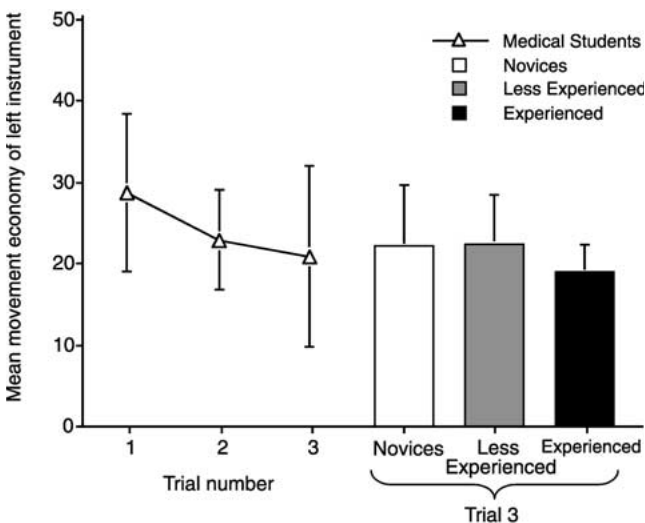
slower than the experienced group ( $F(1,110) = 7.9$ ,  $p = 0.0058$ ), although their times were no different from the inexperienced group or the control group of novices (Fig. 1). There was no difference among the groups in measurements of error (Fig. 2) and economy of movement of the right hand (Fig. 3). In measurements of economy of movement of the left hand, a difference between medical students and the experienced group was found only for trial one ( $F(1,110) = 5.53$ ,  $p = 0.02$ ) (Fig. 4).

The performance of medical students on the economy of diathermy measure was similar to that of the experienced group of surgeons on all trials (Fig. 5). However, the students scored slightly better than the less experienced group of surgeons on all three trials (trial one,  $F = 10.26$ ,  $p = 0.002$ ; trial two,  $F = 9.45$ ,  $p = 0.003$ ; trial three,  $F = 11.22$ ,  $p = 0.001$ ). A similar difference was seen between the medical students and the control group of novices without any medical experience (university students) ([ $dr = 1,110$ ] trial one,  $F = 7.13$ ,  $p = 0.0009$ ; trial two,  $F = 19.3$ ,  $p = 0.0001$ ; trial three,  $F = 15.31$ ,  $p = 0.0002$ ).

If the definition of the criterion level for performance is based on the reproducible performance of the experienced group of surgeons, we can attempt to differen-



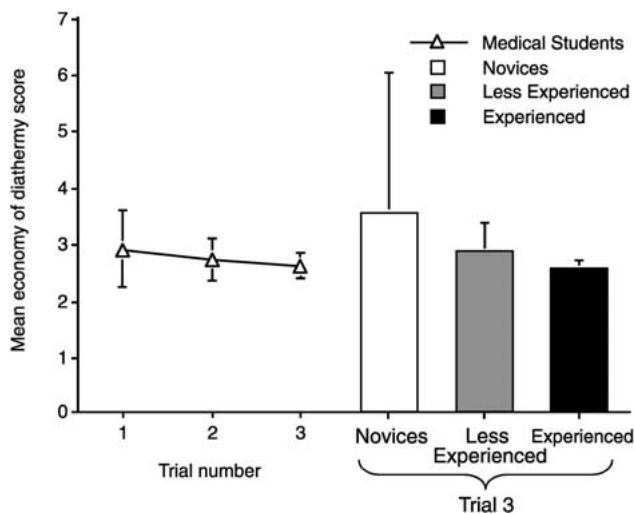
**Fig. 3.** Mean economy of movement Scores ( $\pm$ SD) of the right instrument for the medical students in trials one through three in comparison to the novice, less experienced, and experienced groupson trial three.



**Fig. 4.** Mean economy of movement scores ( $\pm$ SD) of the left instrument for the medical students in trials one through three in comparison to the novice, less experienced, and experienced groups on trial three.

tiate the medical student group into those who may and those who may not have the requisite psychomotor skills to easily obtain laparoscopic expertise. Using the performance of the experienced group of surgeons on trial three as the criterion level, we can stratify the medical student group according to those who fall more than one SD from the mean of the experts and those who fall more than two SD from the mean of the experts (Table 2). The mean of the experienced surgeons can be considered a conservative measure, whereas the mean plus two SD can be considered a liberal measure.

For each measure, 30–62% of medical students did worse than the mean, 12–38% did worse than the mean plus one SD, and 9–27% did worse than the mean plus two SD. When the criterion level was set at the mean of the experienced surgeon, for the measures of time and economy of diathermy, ~30% of medical students did



**Fig. 5.** Mean economy of diathermy scores ( $\pm$ SD) for the medical students in trials one through three in comparison to the novice, less experienced, and experienced groups on trial three.

worse. At the same criterion level, 62% of medical students committed more errors than the mean of the experienced surgeons. If the criteria are made more liberal, we find 38% of medical students making more errors than the expert mean plus one SD and 27% of medical students making more errors than the expert mean plus two SD.

To further analyze the individual performance of the subjects who were performing worse than our most liberal criteria levels on trial three (more than two SD from the experienced surgeons' mean), we looked at improvement between trial one and trial three. This was measured as percentage of improvement between trial one and trial three and was compared to the improvement seen in all subjects (Table 3). The means and SD for both groups across the five measures showed no discernable pattern of improvement. However, for all five MIST-VR metrics, the group as a whole showed a considerably lower SEM.

## Discussion

The data presented here show the results of a large group of laparoscopic neophytes ( $n = 100$ ) trained in three trials on six tasks using the MIST-VR. At the third trial, the mean of the medical students' performance is equivalent to a control group of novices and is not significantly different from the mean of the experienced group of surgeons with respect to error, economy of movement of the right and left hands, or economy of diathermy. There was a statistically significant difference in total time.

We know that the psychomotor performance of experienced surgeons is superior on the MIST-VR, with fewer errors, faster times, and more economical movement of the instruments. They are also more consistent in their performance, as indicated by their smaller SD from the mean scores [8]. It is reasonable to use their

**Table 2.** Percentage of medical students whose scores for trial three were worse than the mean of the experienced surgeons scores, the mean plus one SD, and the mean plus two SD on the five different measures

Measure	Mean	SD	> Mean (%)	> Mean + 1 SD (%)	> Mean + 2 SD (%)
Time	823.2	157.5	31	23	8
Errors	13.8	3.6	62	38	27
Economy of right hand	19.4	3.4	51	32	7
Economy of left hand	19.2	3.6	53	12	12
Economy of diathermy	2.59	0.161	30	16	9

Higher numbers are worse scores

**Table 3.** Mean percentage improvement in scores from trial one to trial three for all subjects tested and subjects who scored more than two SD from the experienced surgeons' mean (criterion level) on trial three on the five measures

	All subjects			Subjects performing worse than criterion level on T3		
	<i>n</i>	Mean improvement T1-T3 (%)	SEM	<i>n</i>	Mean improvement T1-T3 (%)	SEM
Time	100	37.4 (15.05)	1.5	8	31.14 (14.6)	5.17
Errors	100	20.51 (38.42)	3.84	27	3.69 (58.51)	11.26
Economy of left hand	100	15.66 (75.34)	7.53	12	4.9 (53.84)	15.54
Economy of right hand	100	31.46 (24.42)	2.44	7	41.06 (20.56)	7.27
Economy of diathermy	100	8.3 (10.49)	1.05	9	15.44 (7.49)	7.49

T1, trial one; T3, trial three  
SD given in parentheses

performance as a starting point to define benchmarks (criterion levels) for the objective assessment of laparoscopic trainees, although an ideal benchmark would require a much larger sample size of experts.

Using our experts' scores to define performance criteria, we can identify the laparoscopic neophytes who performed worse than our chosen benchmark. As a group, the means of the medical students' performance were not different from those for the experienced groups, but there was greater variability. The greater variability in the medical student group led to substantial numbers of students who performed worse than the criterion levels that we set. The ability of the MIST-VR to discriminate the medical students who performed at the extremes further confirms its role as a useful assessment tool for evaluating the psychomotor component of laparoscopic skills. Furthermore, the MIST-VR may be valuable as a predictor of a subject's potential to acquire the psychomotor skills necessary for laparoscopic surgery.

Laparoscopic surgery requires a definable and unique set of skills that are distinct from those used in open surgery. Certainly, some surgeons are better at these skills than others, and there may be a group of excellent open surgeons who lack the psychomotor, perceptual, or visiospatial ability to become accomplished laparoscopic surgeons. There is ample anecdotal evidence from residency training programs that some residents (and some attending surgeons) just cannot master the skills. As yet, there is no evidence to support these beliefs, aside from subjective evaluations. A reliable, accessible assessment tool that can discriminate psychomotor ability would be a valuable adjunct.

A larger sampling of experts in terms of performance and correlation of MIST-VR and operating room performance would help to define the benchmark criteria

for laparoscopic proficiency. In a randomized double-blind study of MIST-VR training and operating room performance, Seymour et al. defined the training goal for residents undergoing virtual reality training according to the mean of the experts for number of errors and economy of diathermy [16]. All of the residents were successful at meeting the goal, although the number of training sessions varied from six to eighteen. Training according to these performance criteria in the MIST-VR improved completion rates, decreased time, and decreased errors in the operating room. This finding suggests that the MIST-VR is useful in assessing performance during training and that minimum proficiency criteria can be set and evaluated. Our study further suggests that the MIST-VR may be able to identify trainees who lack the psychomotor aptitude to meet proficiency criteria and therefore may not perform as well in the operating room.

One question that this study was not able to address is whether subjects who scored more than two SD from the experienced surgeons' mean would have reached the criterion level if they had been given more than three trials. What can be concluded is that these subjects' performances at the outset of the study were worse than the overall group and that their rates of improvement were about the same as those for the subjects who were performing at criterion level. Whether the poor performers would have eventually reached criterion level is an empirical question that could be answered in another study. However, they did have a considerably larger SEM. This could be a reflection of the smaller number of subjects in this subset, or it could be a reflection of their inconsistent performance. Performance assessment should look not only at the attainment of performance criteria, but also at consistent performance at criterion level [5, 8].

As a simulator, the MIST-VR is relatively simplistic. It lacks haptic feedback, and it re-creates an abstract virtual environment and tasks that are designed to correlate to real-world tasks rather than mimic them. Unlike aviation simulators that incorporate aspects of cognitive function as well as teaching the mechanics of flying, the MIST-VR isolates the psychomotor component of laparoscopy. However, despite its shortcomings, the MIST-VR has been validated as a training and assessment tool by multiple studies, and it has the test-retest reliability  $>0.8$  that is generally required for high-stakes assessment [15, 16, 19]. These data demonstrate that at the novice level the MIST-VR discriminates between individuals with weak vs strong psychomotor abilities, thus completing an important step in validating the MIST-VR as an assessment tool. To develop a comprehensive tool capable of assessing an individual's technical aptitude for laparoscopy, other validated tests for the evaluation of visiospatial and perceptual skills would be needed.

The validation of the MIST-VR shows that even rudimentary virtual reality simulators can be important tools for surgical educators. More complex and realistic simulators are already available, and although a body of evidence comparable to that validating the MIST-VR is currently lacking, it is likely that they too will prove to be effective tools for training in the skills necessary for minimally invasive surgery. Improved virtual environments and haptic feedback are needed to provide the face validity that would make the acceptance of simulation training more palatable to the surgical community; however, the MIST-VR provides a relatively inexpensive, simple system platform and proof of the concept that virtual reality works for assessment and training.

## Conclusion

The MIST-VR can measure psychomotor ability, as well as the variability in performance between subjects with similar experience. When compared to established performance criteria, the subjects can be stratified according to psychomotor ability. This discrimination among levels of technical aptitude may be useful in evaluating and training laparoscopic surgeons.

*Acknowledgments.* This study was funded by a grant to Anthony G. Gallagher from the Economic and Social Research Council (ESRC), UK.

## References

1. Chaudhry 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
2. Chiasson PM, Pace DE, Schlacta CM, Mamazza J, Poulin EC (2002) MIS training in Canada: a survey of general surgery residents. *Surg Endosc* 16(Suppl 1): S202
3. Crothers IR, Gallagher AG, McClure N, James DT, McGuigan J (1999) Experienced laparoscopic surgeons are automated to the “fulcrum effect”: an ergonomic demonstration. *Endoscopy* 31: 365–369
4. Cuschieri A (1995) Whither minimal access surgery: tribulations and expectations. *Am J Surg* 169: 9–19
5. Gallagher AG, Satava RM (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills: learning curves and reliability measures. *Surg Endosc* 16: 1746–1752
6. 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
7. Gallagher AG, McClure N, McGuigan J, Crothers I, Browning J (1999) Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST-VR). *Endoscopy* 31: 310–313
8. Gallagher AG, Richie K, McClure N, McGuigan J (2001) Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. *World J Surg* 25: 1478–1483
9. Hunter JG (2002) The case for fellowship in gastrointestinal and laparoendoscopic surgery. *Surgery* 132: 523–525
10. Jordan JA, Gallagher AG, McGuigan J, McClure N (2000) Randomly alternating image presentation during laparoscopic training leads to faster automation to the “fulcrum effect. *Endoscopy* 32: 317–321
11. Jordan JA, Gallagher AG, McGuigan J, McGlade K, McClure N (2000) A comparison between randomly alternating imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. *Am J Surg* 180: 208–211
12. Jordan JA, Gallagher AG, McGuigan J, McClure N (2001) Virtual reality training leads to faster adaptation to the novel psychomotor restrictions encountered by laparoscopic surgeons. *Surg Endosc* 15: 1080–1084
13. McNatt SS, Smith CD (2001) A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. *Surg Endosc* 15: 1085–1089
14. Ota D, Loftin B, Saito T, Lea R, Keller J (1995) Virtual reality in surgical education. *Comput Biol Med* 25: 127–137
15. Satava RM (1993) Virtual reality surgical simulator: the first steps. *Surg Endosc* 7: 203–205
16. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 236: 458–463; discussion 463–464
17. Smith CD, Farrell TM, McNatt SS, Metreveli RE (2001) Assessing laparoscopic manipulative skills. *Am J Surg* 181: 547–550
18. Surgery Residency Review Committee (2001–2) Surgery resident statistics summary. American Council on Graduate Medical Education, Chicago, IL
19. Torkington J, Smith SG, Rees BI, Darzi A (2001) Skill transfer from virtual reality to a real laparoscopic task. *Surg Endosc* 15: 1076–1079