

VR to OR: A Review of the Evidence that Virtual Reality Simulation Improves Operating Room Performance

Neal E. Seymour

Published online: 3 December 2007
© Société Internationale de Chirurgie 2007

Abstract The use of virtual reality (VR) simulation to train surgeons has been supported by a body of experimental data derived from randomized trials of VR simulation training versus no such training. These investigations have focused on the use of VR devices to train surgeons in laparoscopic and flexible endoscopic skills, and the studies have generally demonstrated that skills acquired through courses of training in VR transfer to the clinical or animal laboratory setting, where assessments of various types have been used to measure performance. These studies, as well as the study model that describes them, and the future of randomized trials of this type are reviewed.

The introduction of simulation methods has had a transformational effect on medical training, and technology-based applications, such as surgical virtual reality (VR), have rapidly progressed from rudimentary concept-demonstrators to practical training modalities. Virtual reality surgical simulation can be defined as the use of a computer to generate an environment with surgical relevance based on mathematical models with which humans can interact by using physical representations of surgical instruments. The essential features of existing VR surgical simulators enable their use to train surgical skills and/or to make inferences about levels of surgical performance. In their latter role, VR training systems are designed to measure

various facets of performance, such as motion and efficiency characteristics, errors, and time to complete a specified task. The systems also make it possible to record this information in a database from which it can be recovered for analysis. The validation of these measurement capabilities has been the subject of intensive study by investigators in several countries, with the aim of ultimately establishing that training in a VR environment improves clinical performance. A number of studies have been undertaken to demonstrate that skills acquired during VR training transfer to the operating room. The background, results, and significance of these studies are reviewed.

Defining skill in simulation

The most significant steps in virtual reality surgical simulator development have been taken in the areas of endoscopic and laparoscopic surgery. Modeling the computer-generated environment and user interfaces for these applications has presented fewer obstacles to developers than those that must be surmounted to simulate open surgical procedures. Furthermore, most of what has been rigorously studied of VR and surgical training effectiveness has been in these areas. At present, VR simulation has been used under both investigative and broad formative training conditions. The latter use presumes that the training is of value, whether or not actual increases in surgeon skill in clinical settings have been demonstrated. The term VR-to-OR, coined by Anthony G. Gallagher, has been used to describe studies designed to show this type of skills transfer. The extent to which VR has been adopted for formative training, and may be adopted in the future, has depended on the quality of such studies.

N. E. Seymour (✉)
Department of Surgery, Baystate Medical Center,
Tufts University School of Medicine, 759 Chestnut Street,
Springfield, Massachusetts 01199, USA
e-mail: neal.seymour@bhs.org

Numerous investigations of the value of surgical VR simulators as tools in performance measurement have been undertaken. Such value signifies that when VR simulators are used to measure performance, the measurements are held to be accurate representations of a user's skill, and that performance data for those who are predictably less skilled (e.g., beginners) and those who are more skilled (e.g., experts) will be accurately scaled to reflect these characteristics. Arguably, this demonstration of *construct* or *contrast* validity for VR surgical simulators is the fundamental requirement for acceptance of a device for performance measurement during formative training. Although this type of validation implies that the performance gap between skilled and unskilled users can be closed by deliberate practice, such a change can only be demonstrated through actual use. In the absence of other information, the performance levels demonstrated by experts are the most appropriate performance targets for novices. This use of specific performance objectives in a proficiency-based training model is held to be the one that is best able to achieve increases in clinical performance [1].

Once performance objectives are determined, there is an opportunity to test the ability of the training subjects to achieve them, and to test the effectiveness of completed training based on some measurement of relevant real-world performance. The manner in which learners are exposed to VR in assessment and training can influence the rate at which performance curves for novices and experts converge. Factors that might influence achievement of performance (and hence educational objectives) include use of mentoring and real-time feedback during training, as well as the difficulty of the task being trained. The latter factor might reflect some fundamental challenge of the task (e.g., complexity), or software configuration assignments pertaining to difficulty. Examples of the latter include tolerance for registering an error event or inability to progress in the task without achieving a specific intermediate step (e.g., needle entry into simulated tissue not permitted unless proper needle position in instrument achieved). These characteristics underscore some of the fundamental differences between VR and non-VR training systems.

VR to OR: demonstrating transfer of skill

The actual demonstration of skills transfer necessitates use of tools to accurately measure the clinical skills relevant to the VR training. In connection with this, the performance measurement capabilities of the simulator are needed to define the specific levels of skill in simulation that are predictive of corresponding levels of clinical skill. Without performance data in both settings, the simulator cannot be used to infer skill or a lack of skill. Predictive validity

testing pertains to the demonstration that measured performance in the simulation predicts future measured performance in the real-world activity being simulated. The demonstration of predictive validity is the basis for the use of simulation performance measurements either to make decisions regarding competence or to set requirements for advancement. This must be distinguished from a study intended to demonstrate that skill acquired in simulation transfers to the clinical environment.

To demonstrate that skills gained during simulation training transfer to the operating room, a measurable end point of training that maximizes the likelihood of seeing such an effect should be defined. For the most part, studies that have sought to demonstrate transferability of VR-acquired skills have done so at the final phase of a course of training in VR that uses proficiency-based [2, 3], repetition-based [4, 5], and time-based [6–8] training models. All cited studies have used drug trial-like study designs, where subjects have been randomized to VR training and control study groups. There are no widely accepted norms for definition of an adequate control group, however, and it must be appreciated that control group characteristics might profoundly affect the results of such a study. For example, a control group that receives no specific training might be expected to perform differently from one that received some form of traditional training as an alternative to VR. The power of inclusion of a control group into the study design lies in the ability to establish a cause-effect

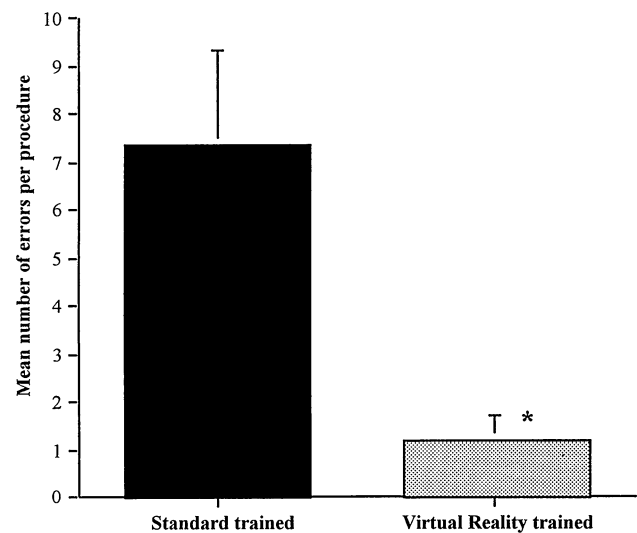


Fig. 1 Results of surgical resident performance after virtual reality (VR) training of skills deemed vital to a specific clinical task (excision of the gallbladder from the liver). These results demonstrate the strength of the randomized trial in defining the skills transfer effect by virtue of the control group (“Standard trained”) that had not received any systematic training aside from their normal clinical duties (from *Annals of Surgery* 2002;236:458–464, reproduced with permission.)

relationship between the simulator training and improvement in clinical performance (Fig. 1).

One of the more significant difficulties presented by this model of skills transfer testing is the requirement for valid measures of clinical performance. Because most VR simulation devices provide a number of part-task training experiences, it is necessary to select a clinical assessment methodology that examines specific skills that might imparted with such devices. In the absence of an appropriate pre-existing method, one must be designed. It is vital not to overreach the capabilities of the VR system by attempting to measure clinical performance in areas that are beyond those trained by the VR task.

Conditions in the clinical operating room impose practical limits on performance measurement methods that might employ obtrusive equipment or personnel that can more easily be used in the training laboratory. In addition, some types of skills are not appropriate to test in the clinical OR in consideration of patient safety. This aspect of study design is influenced by the makeup of the subject group (e.g., medical students versus residents), as well as the nature of the task being assessed. Some investigators have addressed these issues by using animal models that recreate some but not all of the conditions of the clinical OR. A variety of means of assessing operative performance have been described, including assessment of video recordings [9, 10], global assessment methodologies applied to live observations, or videos [11, 12]. In addition, various designed-for-purpose operative performance assessment methods that employ motion analysis [13, 14] have been designed, as have more complex methods such as those that use hidden Markov models [15]. The level of validity testing to which these methods have been subjected adds another level of complexity to the demonstration of skills transfer from VR. To date, skills transfer studies that examine operating room performance have used direct observation and video assessments.

Recent reviews of published skills transfer studies have concluded that this body of data is generally supportive of the use of VR simulation devices to improve operative skills [16, 17]. Haque and Srinivasan recently reported the examination of skills transfer from VR by meta-analysis of task completion time and error scores data for six studies reflecting a combination of laparoscopic and flexible endoscopic VR training [16]. This suggested a strong educational effect on skills transfer, and it is the only study of this type to attempt to show effectiveness of VR training on a broad basis. It remains apparent, however, that individual attempts to demonstrate skills transfer from VR have been quite limited, as reflected in the relatively small number of publications and the limited scope of the skills examined [2–8].

Although the number of surgical simulation developers and devices has grown [18], only a limited number of

laparoscopic and flexible endoscopic VR training devices have been studied specifically for skills transfer characteristics. Most of these studies have defined time-based criteria for VR training completion or have required performance of a fixed number of repetitions prior to testing of clinical skills (Table 1). The use of time (length of time or number of sessions over which training occurred) as a training endpoint allows for potentially large variations in both exposure to the VR activity and in subject skill level. Proficiency-based training has been employed less frequently [2, 3].

Of the seven published studies of laparoscopic skills transfer that were identified, one failed to demonstrate transfer of skills [6]. Provided the study subject groups are homogeneous from the standpoint of pretraining performance, this finding suggests that either the training system is deficient (e.g., it does not increase the skill level of the learner) or the clinical assessment methodology did not permit identification of the training effect. Unless there is more detailed information about the training available, the specific confounding issue is predictably difficult to identify. Training and control group homogeneity can be shown by the use of pretraining psychometric tests (including the use of the simulator as a psychometric study tool) [2], or a pretraining clinical assessment of all study subjects [3]. The latter assessment has the potential disadvantage of constituting a training experience that might affect both control and study subjects and thus would dilute the ability to demonstrate a simulator training effect. The cited study that failed to identify skills transfer may have been hampered by (1) a fairly complex “clinical” task (suture loop application to a loop of bowel prepped to have the appearance of an appendix in an anesthetized swine model) performed by medical student study subjects after only a brief tutorial, and (2) a very brief course of training (basic manipulative tasks on MIST-VR) that was probably not well matched to this procedure. It is important to recognize that the inability to demonstrate a skills transfer effect does not necessarily signify that the training or the simulator is without value. Rather, the study results should prompt examination of the details of implementation of VR training and clinical skills assessment in order to better understand how to use or improve these tools.

Several skills transfer studies of VR flexible endoscopic trainers have been conducted [19–25], and the results have generally been similar to those reported for laparoscopic VR training. Global measures of performance, time to procedure completion, and achievement of specific procedural goals have been used as clinical skill metrics. However, some of these studies have also examined clinical outcomes, including patient discomfort and satisfaction, as metrics for effectiveness of training [22–24]. Efforts to look beyond the technical aspects of the

Table 1 Published reports of randomized trials of virtual reality training and skills transfer effects

Study	Study type/training model	Platform(s)	Clinical skill/assessment method	Results
Ahlberg et al. [6]	Randomized trial/time-based training	MIST-VR ^a	Laparoscopic appendectomy (pig)/blinded rater video analysis	Medical student performance with training was no better than that of non-trained controls
Hyltander et al. [8]	Randomized trial/undefined	LapSim ^b	Various laparoscopic skills (pig)/blinded rater video analysis	Medical student performance improved with training
Hamilton et al. [7]	Randomized trial/time-based training	MIST-VR ^a	Laparoscopic cholecystectomy (human)/blinded rater global assessment	Surgical resident performance improved with training
Seymour et al. [2]	Randomized trial/proficiency-based training	MIST-VR ^a	Laparoscopic cholecystectomy (human)/blinded rater video analysis	Surgical resident performance improved with training
Grantcharov et al. [4]	Randomized trial/number of repetitions	MIST-VR ^a	Laparoscopic cholecystectomy (human)/blinded rater video analysis	Surgical resident performance improved with training
Youngblood et al. [5]	Randomized trial/number of repetitions	LapSim ^b	Various laparoscopic skills (pig)/blinded rater video analysis	Medical student performance improved with training
Ganai et al. [3]	Randomized trial/proficiency-based training	Endotower ^c	Angled laparoscope (pig)/blinded rater assessment	Medical student performance improved with training
Ost et al. [19]	Randomized trial/number of repetitions	Accutouch ^d	Bronchoscopy (human)/direct observation	Fellows performed task faster and more skilfully after VR training than untrained controls.
Rowe et al. [20]	Randomized trial/undefined	Accutouch ^d	Fiberoptic intubation (human)/blinded rater video analysis	Residents performed task faster and more skilfully after VR training than untrained controls
Gerson and Van Dam [21]	Randomized trial/number of repetitions	Immersion sigmoidoscopy trainer ^d	Sigmoidoscopy (human)/direct observation and patient survey	Residents performed less skilfully after VR training than a cohort trained on actual clinical procedures
Sedlack and Kolars [22]	Randomized trial/time-based training	Accutouch ^d	Colonoscopy (human)/direct observation and patient survey	VR trained resident-performed procedures were associated with less patient-reported discomfort
Sedlack et al. [23]	Randomized trial/time-based training	Accutouch ^d	Colonoscopy (human)/direct observation and patient survey	GI Fellows performed more skilfully after VR training
Blum et al. [24]	Randomized trial/time-based training	Accutouch ^d	Bronchoscopy (human)/direct observation	Residents performed more skilfully after VR training
Knoll et al. [25]	Nonrandomized trial/undefined	UroMentor ^e	Ureterorenoscopy (human)/undefined	Urology residents performed faster after VR training than a nontrained cohort

^a Mentice AB, Göteborg, Sweden^b Surgical Science, Göteborg, Sweden^c Verifi Technologies, Inc., Elizabethtown, PA^d Immersion Medical, Gaithersburg, MD^e Symbionix Ltd., Lod, Israel

clinical task are very important, and improved outcomes will ultimately provide the most compelling argument for the use of VR simulation training. From the cited studies it remains to be defined how the simulator-trained behaviors may have affected patient comfort, or whether the use of analgesic or sedative medications differed in either the VR-trained group or the non-trained group.

Although most of the study designs paralleled those for laparoscopic skills transfer studies in their use of VR-trained and non-trained study groups, all of these studies employed non-blinded direct observations for clinical assessments, raising observer bias concerns. One of the studies [22] used a control group that received training on 10 actual clinical flexible sigmoidoscopies prior to clinical performance testing on the same task. Not unexpectedly, these subjects performed better than a group trained exclusively on a VR simulator prior to the same clinical assessment. It is difficult to draw specific conclusions from this type of comparison both because it does not define the training benefit of the VR device or support the most appropriate application of a lab-based device (e.g., use of VR simulation training as a preparatory step to actual clinical endoscopies on humans). A more useful comparison might be between a VR simulator and non-VR training device, such as a benchtop anatomic model.

Current views of VR-to-OR studies

Although most of the cited VR-to-OR studies show skills transfer effects, it is necessary to look beyond these positive data, and to formulate practical recommendations pertaining to formative training of operative skills. The VR-to-OR skills transfer study model should be viewed as a means of demonstrating the value of a very deliberately designed VR training activity, rather than of the simulator itself. Because the scope of these studies has been rather narrow, surgery educators examining the prospect of adding VR simulators to their training programs currently face some uncertainties with regard to implementation of effective VR training activities. Only three skills transfer studies involving laparoscopic surgery have used surgical residents as study subjects. The selection process leading to this subject pool makes it substantially different from one comprised of medical students, and extrapolating training results from studies examining medical student skills to resident training may be problematic. There is an imperative to achieve the highest possible level of performance in a surgical trainee with lab-based training because of the potential implications for patient care. This means deliberate selection of VR task(s), selection of task difficulty levels, selection of duration of training, and definition of reasonable performance objectives.

Optimally, performance objectives should be selected and vetted to be the skills of experts, and training protocols for residents should be designed to allow achievement of those skills without restrictions that might be imposed by a time limit such as a medical student's surgical rotation length. Skills transfer from VR for surgical residents should be meaningful both in scale and in comparison to the performance expected from experts. This requires assessment of both resident and expert clinical performance on tasks that may be inappropriate for a medical student to perform even with preliminary training.

To date, none of the reports on laparoscopic skills transfer from VR contrast post-training clinical skills to those of expert surgeons in the clinical OR. The VR-to-OR study design with non-trained study subjects is not required to accomplish this. In 2005, the European Association of Endoscopic Surgeons (EAES) published consensus guidelines on validation of VR surgical simulators [17]. This work forwarded levels of recommendation for VR systems based on specific evidence-linked criteria for demonstration of construct and concurrent or predictive validity. These criteria define randomized trials as constituting the highest qualitative level of evidence, warranting the highest levels of recommendation. This view of the scientific rigor to which systems must be subjected in order to be considered "valid" best represents current thinking and planning for the use of VR in any decision-making process intended to certify surgeon competence. However, as stated previously, the concept that the simulator is validated by such studies ought to be dropped in favor of the view that the training curriculum using the simulator is shown to be effective or ineffective. The current focus on simulator validity does not necessarily follow models of simulator use in other undertakings where long-standing developmental and implementation experience has permitted firm assumptions in support of use of simulation as a bedrock tool to train skills and to answer performance-related questions. An alternative use of randomized trials to show skills transfer with VR training might be to compare training curricula with the aim of maximizing training benefit.

Discussion

At this stage in the use of VR surgical training platforms, ethical questions may arise concerning the VR-to-OR model described above. A major concern is the use of "no training" control subjects, particularly if the studies involve application of clinical skills in human patients. As we move forward to an era where high-fidelity VR procedural simulation becomes feasible, ethical questions become more relevant. In order to take advantage of the advancing capabilities of VR simulation, a new phase of study and

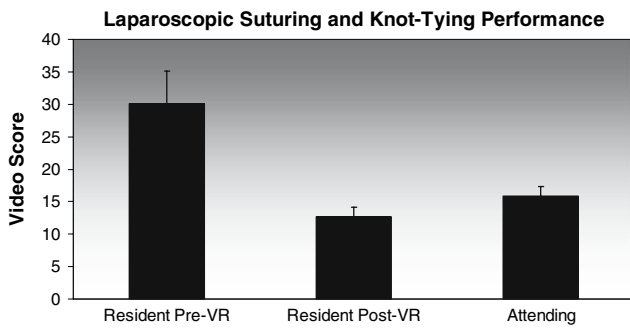


Fig. 2 Blinded rater video analysis of surgical resident performance of intracorporeal laparoscopic suturing and knot-tying before and after VR training of this task. Reference to resident case logs defined a minimal exposure to clinical experiences that might have contributed to improved performance over the course of training. The educational goal of expert performance achievement was realized for the group, but on an individual basis, some residents can be defined as requiring additional training to achieve this level of performance. The disadvantage of failure to use an untrained control group is an inability to assess the training effect of the initial assessment

validation ought to be envisioned that defines additional methods to examine outcomes of training. Alternatives to randomized trials of the type characterized as “VR-to-OR” would examine results of actual usage within the framework of well-designed curricula, rather than small-scale investigative efforts to provide ongoing performance data and to establish training effectiveness. Concurrent and predictive validity study models require multiple methods of assessments to establish correlation of contemporaneous and future lab and clinical performance with performance achieved in VR. These types of data can be obtained under non-investigative conditions during the course of actual training. One example of concurrent validity might be correlation of performance in VR with contemporaneous performance in a gold standard training lab test such as an objective structured assessment of technical skill (OSATS) evaluation [26]. Although predictive validity, or the ability to predict future clinical performance based on measured performance in VR, is sometimes associated with the VR-to-OR study model, the demonstration that an assessment system is predictively valid does not require untrained control subjects. Predictive validity implies a higher level of fidelity to the real situation in which the task is used than concurrent validity, and it can also be examined on an ongoing basis in the course of actual formative training and clinical activities. This would require development and routine use of assessment systems for clinical performance that are not widely used today. Similarly, serial tests of lab and clinical performance (repeated measures study model) can suggest a cause–effect relationship between VR simulator training and improved operative performance [27], with effectiveness defined by achievement of measurable expert performance goals (Fig. 2).

Virtual reality training systems are intended to create new experiential learning opportunities that can serve as safe and effective alternatives to more traditional learning venues, such as the clinical operating room. The optimal use of new VR training platforms requires that the best possible assessment methods for the clinical OR be devised and validated. More intensive evaluation of this type could be used to guide implementation of innovative training methods such as VR, based on a dynamic process of continuous examination of performance, identification of performance outliers, and modeling of training activities to achieve carefully selected training goals based on expert performance behaviors.

References

- Gallagher AG, Ritter EM, Champion H, et al. (2005) virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 241:364–372
- Seymour NE, Gallagher AG, Roman SA, et al. (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 236:458–464
- Ganai S, Donroe J, St. Louis M, et al. (2007) Virtual reality training improves angled laparoscopic telescope skills in novice laparoscopists. *Am J Surg* 193:260–265
- Grantcharov TP, Kristiansen VB, Bendix J, et al. (2004) Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg* 91:146–150
- Youngblood PL, Srivastava S, Curet M, et al. (2005) Comparison of training on two laparoscopic simulators and assessment of skills transfer to surgical performance. *J Am Coll Surg* 200:546–551
- Ahlberg G, Heikkinen T, Iselius L, et al. (2002) Does training in a virtual reality simulator improve surgical performance? *Surg Endosc* 16:126–129
- Hamilton EC, Scott DJ, Fleming JB, et al. (2002) Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. *Surg Endosc* 16:406–411
- Hyltander A, Liljegren E, Rhodin PH, et al. (2002) The transfer of basic skills learned in a laparoscopic simulator to the operating room. *Surg Endosc* 16:1324–1328
- Eubanks TR, Clements RH, Pohl D, et al. (1999) An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg* 189:566–574
- Seymour NE, Gallagher AG, Roman SA, et al. (2004) Analysis of errors in laparoscopic surgical procedures: a new methodology. *Surg. Endosc* 18:592–595
- Vassiliou MC, Feldman LS, Andrew CG, et al. (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 190:107–113
- Gosman GG, Simhan HN, Guido RS, et al. (2005) Focused assessment of surgical performance: difficulty with faculty compliance. *Am J Obstet Gynecol* 193:1811–1816
- Datta V, Bann S, Mandalia M, et al. (2006) The surgical efficiency score: a feasible, reliable, and valid method of skills assessment. *Am J Surg* 192:372–378
- Datta V, Mackay S, Mandalia M, et al. (2001) The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg* 193:479–485

15. McBeth PB, Hodgson AJ, Nagy AG, et al. (2002) Quantitative methodology of evaluating surgeon performance in laparoscopic surgery. *Stud Health Technol Inform* 85:280–286
16. Haque S, Srinivasan S (2006) A meta-analysis of the training effectiveness of virtual reality simulators. *IEEE Trans on Information Technol Biomed* 10:51–58
17. Carter FJ, Schijven MP, Aggarwal R, et al. (2005) Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc* 19:1523–1532
18. Dunkin B, Adrales GL, Apelgren K, et al. Surgical simulation: a current review. *Surg Endosc* 21:357–366
19. Ost D, DeRosiers A, Britt EJ, et al. (2001) Assessment of a bronchoscopy simulator. *Am J Respir Crit Care Med* 164:2248–2255
20. Rowe R, Cohen RA (2002) An evaluation of a virtual reality airway simulator. *Anesth Analg* 95(1):62–66
21. Gerson LB, Van Dam J (2003) A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. *Endoscopy* 35:569–575
22. Sedlack RE, Kolars JC (2004) Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. *Am J Gastroenterol* 99:33–37
23. Sedlack RE, Kolars JC, Alexander JA (2004) Computer simulation training enhances patient comfort during endoscopy. *Clin Gastroenterol Hepatol* 2:348–352
24. Blum MG, Powers TW, Sundaresan S (2004) Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. *Ann Thorac Surg* 78:287–291
25. Knoll T, Trojan L, Haecker A, et al. (2005) Validation of computer-based training in ureterorenoscopy. *BJU Int* 95:1276–1279
26. Martin JA, Regehr G, Reznick R, et al. (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 84:273–278
27. Earle DB, Thompson RE, Kuhn JN, et al. (2006) The use of a virtual reality trainer to increase surgical resident skill on complex laparoscopic tasks. *Surg Endosc* 20:1:S342