SURGICAL ENDOSCOPY
and Other Interventional Techniques

The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance

M. P. Schijven, J. Jakimowicz

Department of Surgery, Catharina Hospital Eindhoven, Michelangelolaan 2, P.O. Box 1350, 5602 ZA Eindhoven, The Netherlands

Abstract

Background: This study assesses the acquisition of skill and the learning curve associated with the performance of the clip-and-cut task on the Xitact LS 500 virtual reality (VR) simulator in laparoscopic cholecystectomy.
Methods: A group of 33 residents and interns with no previous laparoscopic experience participated in the study. All participants received a 1-h familiarization tour on the simulator. Thirty participants completed a full course of 30 simulation runs over 3 days (10 runs per day). The outcome parameters were a previously validated sum-score and time to complete performance.
Results: Group demographics were similar. Of the participants who completed the full study, 16.7% appeared to have such a high level of innate psychomotor abilities that they were considered proficient in the task immediately after the initial familiarization tour. Most participants (63.3%) had a moderate level of innate abilities, and their performance improved through repetitive VR training. In our study, 20% of the participants had such a low level of innate abilities that they were unable to achieve an acceptable performance in our minimal-access surgery (MAS) simulation.
Conclusions: Learning curves cannot be assessed by examining the repetitive training of only one person. There seem to be four different performance profiles, reflecting the fact that some people are more adept than others to be trained by MAS procedural VR simulation. For participants receptive to training—63.3% in this study—proficiency in the task occurs after ~25 simulative runs.

Key words: Virtual reality — Minimal-access surgery — Simulation — Learning curve — Validation — Surgical training

Improvement in the performance of surgery depends, in fact, on practicing surgery [2]. Tight operating room schedules, shortened training curriculum for residents and medicolegal issues have led surgical educators to question whether the operating theater should be the primary teaching environment for the initial acquisition of surgical skill [8, 17]. This dilemma points to the need for supplementary and innovative approaches for the teaching of surgical skills outside the operating theater. Currently, a 3-day hands-on basic surgical skills course is a mandatory part of training in surgery in the U.K. [10]. Indeed, the directors of many surgical residency programs now recognize the need for additional training programs to supplement the more traditional apprenticeship graded responsibility model of surgical training.

The actual performance of surgical tasks is known to improve with experience—that is, through standardized repetition. Improvement tends to be most rapid at first; subsequently, it tails off over time until a steady state of performance is reached [14]. The term "learning curve" is often used to describe this phenomenon. In fact, the learning curve is the graphic representation of the relationship between experience with a procedure and an outcome variable, such as operative time or complication rate [21]. Learning curves have long been recognized in areas other than health care assessment, such as psychology, manufacturing, and aviation. The term was first used in health care in the 1970s and came to greater prominence in the 1980s after the introduction of MAS.

MAS places specific strains on the surgeon, requiring particular psychomotor abilities and skill to overcome difficulties imposed by the videoscopic surgical interface. Studies show that laparoscopic surgery is associated with a higher rate of complications than open surgery [5, 13]. With the widespread application of laparoscopic techniques in the early 1990s, an associated two- to three-fold increase in bile duct injuries was documented among certified surgeons. The risk of complications is known to be highest during the early part of the surgeon's MAS experience [19]. Historical estimates of the number of cases needed to master the procedure of

Correspondence to: M. P. Schijven

laparoscopic cholecystectomy range from eight- to 40. A study by the Southern Surgeons Club that used multivariate regression to analyze $> 8,000$ procedures showed that $> 90\%$ of all bile duct injuries occur within the first 30 cases performed by an individual surgeon [12].

When performing MAS, trainee surgeons cannot easily mimic their mentor's actions or maneuvers without actually manipulating the laparoscopic instruments in an initially disorienting, two-dimensional environment [11]. It is known that surgical skills—and in particular the complex psychomotor skills needed for endoscopic surgery—are in part innate and in part learned through extensive repetitive practice of a procedure [18]. Recent advances in virtual reality (VR) technology have led to the development of VR surgical skills simulators. These novelties are promising assets for the training and assessment of surgical skills [1, 7]. VR simulators provide an opportunity for repetitive practice, allowing for trial and error in the acquisition of new skills without the pressures or consequences of clinical reality. Furthermore, simulators afford flexibility and independence, because training does not depend on the presence of an instructor. Furthermore, simulators offer excellent opportunities for the assessment of surgical skills. The very nature of laparoscopic surgery, with its videoscopic interface, makes it likely to benefit from developments in VR [3].

Ultimately, learning curves for laparoscopic surgery could be shortened by achieving a good and stable performance on the simulator, leading to proficiency in the in vivo procedure. Our study focuses on the acquisition of skill and the learning curve associated with the task-oriented clip-and-cut scenario of the laparoscopic cholecystectomy, as represented by laparoscopic the Xitact LS 500 cholecystectomy VR simulator.

Materials and methods

Subjects

Thirty-three hospital residents and last-year interns, all without any laparoscopic experience, participated in the study. Participants received a 1-h familiarization protocol on the simulator, introducing them to the laparoscopic cholecystectomy clip-and-cut scenario. Participants followed a step-by-step teaching schedule for the laparoscopic cholecystectomy clip-and-cut task. This teaching schedule incorporates live video clips of the simulated in vivo and the VR clip-and-cut procedure, a color-guided teaching approach showing the exact area and preferred sequence for the placement of the clips on the virtual cystic duct and artery, specific instruction on what are regarded as common faults and/or resulting problems, and finally a free-form exercise without color guidance. Feedback, through an assessment sheet, showing the end result of the procedure, and via comments from the instructor, was given after the 1-h familiarization. Each participant performed the procedure 30 times—that is, 10 consecutive procedures per session over 3 consecutive days.

Simulator

The Xitact LS 500 laparoscopic cholecystectomy simulator is a modular VR training platform that was developed for training and education in a variety of laparoscopic skills. It is a hybrid simulator that combines a physical object (the Optable, or "virtual abdomen")

Fig. 1. The Xitact LS 500 laparoscopic cholecystectomy simulator.

with a computer software simulation providing the visual image and tactile feedback. The Xitact incorporates basic surgical skills, the clip-and-cut task for the laparoscopic cholecystectomy, and a peritoneal dissection module for opening Calot's triangle. The module used in this study for determining learning curves was the clip-and-cut task, because this module has been previously validated [16, 17]. The Xitact LS 500 was developed and is registered by Xitact SA (Morges, Switzerland) (Fig. 1).

Statistical analysis

The measurement of the performance of a new task—e.g., the learning curve—poses specific difficulties. Identification of the correct parameter to measure performance is one of them. In MAS, one could opt for the most obvious measurements, such as complications (bleeding or leakage) or ultimately conversion. Nevertheless, they may be too infrequent or not useful (because they tend to be dichotomous) for statistical analysis. Time is considered to be a parameter of importance, although not necessarily a reflection of proper outcome. In our view, statistical methods exploring the learning curve should address three basic aspects of the performance: the aspect of *learning* itself (was there change in performance over time?), the aspect of *proficiency* (when is there no further change in performance; e.g., what is the asymptote of the learning curve?), and the aspect of *stability* (is the change in performance stable?). Parameters of interest must therefore be continuous in outcome. Performance in this study was expressed through a previously validated Xitact-specific performance score [17].

Curve estimation, fitting a line to individual performance data, offering least squares regression was used for trend analysis. Lines reflecting the observed performance outcome and the best-fitting statistical model were plotted. For comparison of the demographics, the Kruskal-Wallis test was chosen. The Statistical Package for the Social Sciences (SPSS) ver. 9.0 (SPSS Inc., Chicago, IL, USA) was used for statistical calculations.

Table 1. Group demographics vs test statistics

	Sex	Age	Dexterity	Speciality	Year of training
Chi-square	3.805	1.347	6.429	1.804	5.067
df	3	3	3	3	3
Asymp. sig.	0.283	0.718	0.093	0.614	0.167

df, degrees of freedom
[a] Kruskal-Wallis test
[b] Grouping variable: GROUP
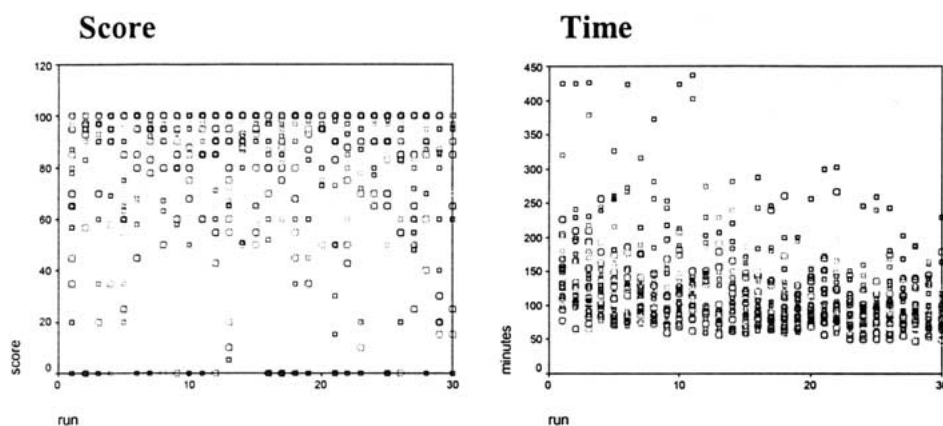
Fig. 2. Scores and times to complete each run for each participant in all groups.
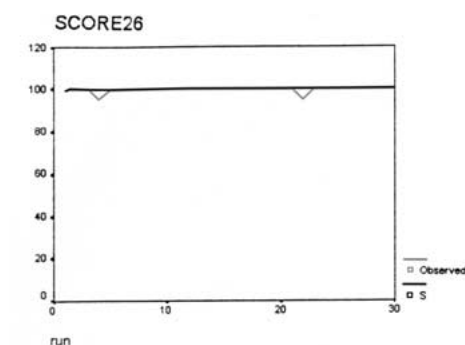
participant nr. 26

Fig. 3. Group 1. High level of innate abilities, gaining little extra improvement through virtual reality training.

Results

Demographics

The mean age of the participants was 28 years (range, 21–35). There were 18 men and 15 women in the study. Nineteen were right-handed, two were left-handed, and two were ambidextrous. Eleven participants were interns. Two were residents in emergency medicine, six were residents in radiology, three were residents in urology, one was a resident in cardiology, three were residents in pulmonology, two were residents in anesthesiology, and five were residents in internal medicine. Three participants could not complete the required 30 runs. They were therefore excluded from further analysis. When Kruskal-Wallis nonparametric tests were used for comparison of the four groups, there were no significant differences (Table 1).

Plots 1

All groups. When scores and time to complete each run for each participant were plotted in one overall graph, there seemed to be no correlation at all (Fig. 2). Scores are dispersed evenly throughout the three sessions of 10 runs. Time needed to complete the runs is also distributed in an incoherent manner with no apparent decrease in time over the 30 runs. This lack of coherence could only be explained by assuming that there were different "sets" of profiles that made interpretation of the results impossible when they were displayed together in one graphic plot. Therefore, individual curve estimations were created to identify the different sets of profiles.

Curve estimation. There appeared to be four different types of curves. Different models (linear, logarithmic, power, and S) were used to estimate the best-fitting curve. The S model—$Y = e^{**[b_0 + [b1/t]]}$—appeared to have the best overall fit. For each group, a representative curve of one of the participants was chosen as an example.

Group 1 was labeled "High level of innate abilities, gaining little extra improvement through VR training" (Fig. 3 and Table 2). Neither the linear model nor the S curve were able to explain the variance in the scores in this group, because there is little variation. For this participant (no. 26), Rsq linear is 0.006, Rsq S is 0.000 (NS).

Table 2. Classification of groups

Group	Profile	Participants	%
1	High level of innate abilities, gaining little extra improvement through VR training • No score < 50 in any run • No score < 80 during the last series of 10 runs • Standard deviation over all runs ≤ 10 • Overall mean ≥ 90	16, 19, 21, 22, 26	16.7
2	Moderate level of innate abilities, gaining improvement and stability through VR training • No scores of 0 in the last two series of 10 runs • No scores < 50 in the last series of 10 runs • Standard deviation over all runs $10 \leq 30$	4, 5, 7, 8, 9, 17, 24, 31, 33	30.0
3	Moderate level of innate abilities, gaining unstable improvement through VR training • No scores of 0 in the last series of 10 runs • Standard deviations over all runs $30 \leq 40$	1, 3, 6, 11, 15, 20, 23, 28, 29, 32	33.3
4	Low level of innate abilities, not gaining improvement through VR training • • Scores of 0 in last series of 10 runs • Standard deviations over all runs > 40	10, 12, 13, 14, 27, 30	20

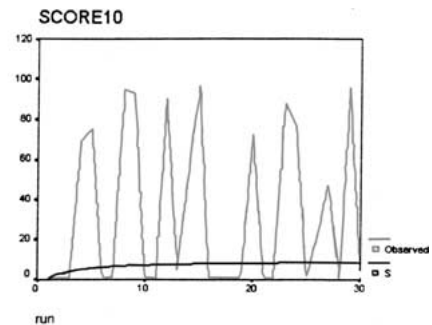VR, virtual reality
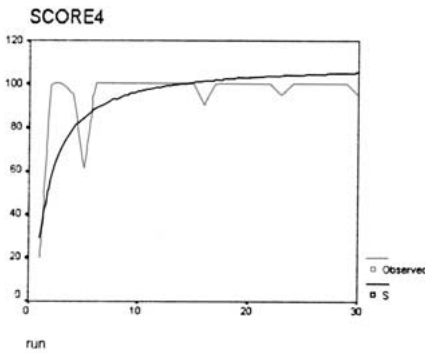
participant nr. 4

Fig. 4. Group 2. Moderate level of innate abilities, gaining improvement and stability through virtual reality training.
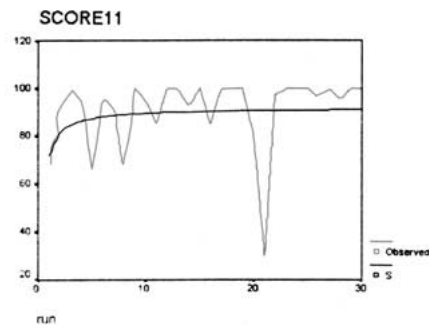
participant nr. 11

Fig. 5. Group 3. Moderate level of innate abilities, gaining unstable improvement through virtual reality training.

participant nr. 10

Fig. 6. Group 4. Low level of innate abilities, not gaining improvement through virtual reality training.

Group 3 was labeled "Moderate level of innate abilities, gaining unstable improvement through VR training" (Fig. 5 and Table 2). The S curve cannot be used explain the variance in scores in this group. Much variation is present throughout the runs, and some participants do better in their second series of runs than in their third series. For this participant (no. 11), Rsq S is 0.038 (NS).

Group 4 was labeled "Low level of innate abilities, not gaining improvement through VR training" (Fig. 6 and Table 2). The S curve cannot explain the variance in the scores in this group. Much and large variations are present throughout the runs, and participants are unstable in their performance throughout the runs. For this participant (no. 10), Rsq S is 0.043 (NS).

Plots 2

When the scores were plotted for group 1 time to complete each run and score vs time, there appeared to be little dispersion (Fig. 7). There seems to be a strong relation between time and score, indicating that the runs with the lower scores actually take longer to complete. Also, participants become faster over runs while preserving their high scores.

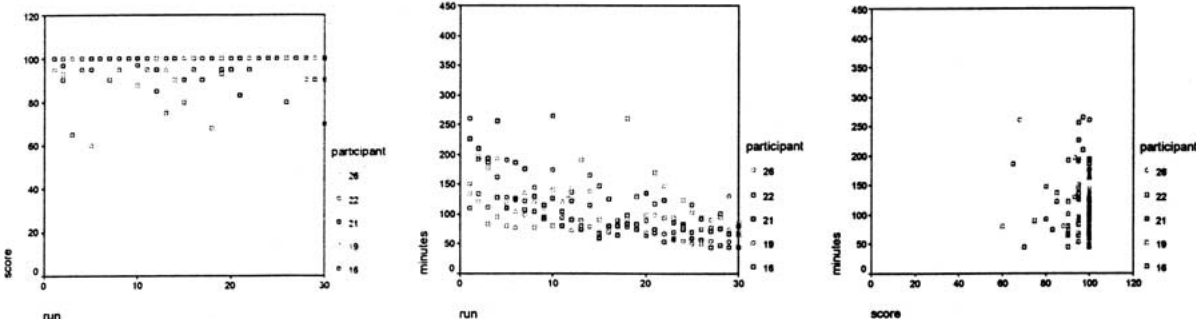Group 2 was labeled "Moderate level of innate abilities, gaining improvement and stability and through VR training" (Fig. 4 and Table 2). The S curve can be used to explain the variance in the scores in this group, because there is a definite learning curve. For this participant (no. 4) Rsq S is 0.694, Sigf is 0.000. That is, 69% of the observed variance is explained by the model.

Score / Time / Score vs Time (Fig. 7)

Fig. 7. Group 1. High level of innate abilities, gaining little extra improvement through virtual reality training.
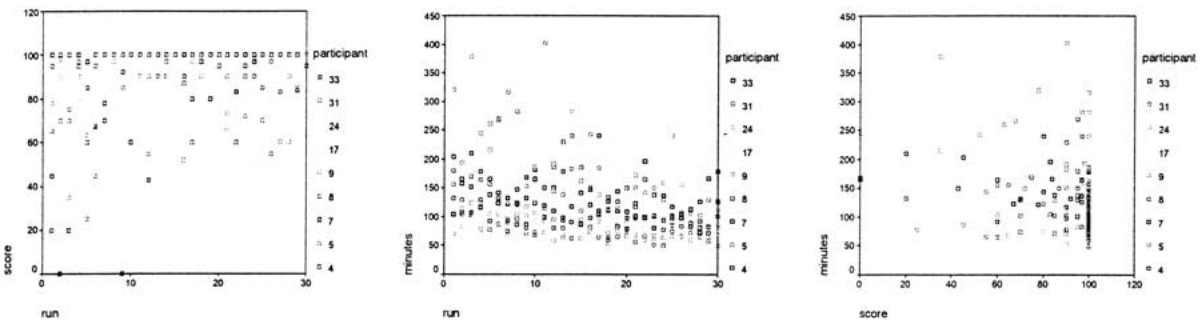

Fig. 8. Group 2. Moderate level of innate abilities, gaining improvement and stability through virtual reality training.

When scores were plotted for group 2, time to complete each run and score vs time, there appeared to be some dispersion, especially in the first series of runs (Fig. 8). There seems to be a relation between time and score, indicating that the runs with the lower scores actually take longer to complete. However, perfect runs sometimes take more time to complete, whereas nonperfect runs may take less time.

When scores were plotted for group 3, time to complete each run and score vs time, appeared to be large dispersions throughout runs (Fig. 9). There seems to be little relation between time and score.

When scores were plotted for group 4, time to complete each run and score vs time, there appeared to be very large dispersions throughout runs (Fig. 10). There seems to be no relation between time and score.

If we look specifically at the curves of the participants in groups 2 and 3, the average number of runs needed to be in the "safe zone" that is, to complete the procedure without harming the common bile duct or incurring other troublesome errors—seems to be ~25 runs.

Discussion

Surgical competence has two major components. The first one is cognitive competence, which requires cognitions mainly based on surgical, anatomical, and medical knowledge. The second one is technical skill in surgery, which is the result of a person's innate abilities to perform a specific surgical task combined with repetitive training in that procedure. There are numerous theories that attempt to explain the acquisition of technical skill. In combination, they emphasize the importance of modeling, repetitive practice, and formative feedback. Obviously, both components are essential for a surgeon to become competent. Assessment tools for measuring cognitive competence are widely available; it is the technical aspect that suffers from poor and subjective assessment strategies [21]. VR studies, so far have lacked the power to measure fundamental issues in the learning of motor skills adequately [6]. In response to the growing need for better methods to assess surgical competence, an initial framework has recently been devised to standardize definitions, measurements, and criteria for objective metrical assessment [15]. There is general agreement that, as with any method of teaching, VR simulators should be evaluated repeatedly before they widespread implementation into surgical education [4]. No doubt, this principle is even more important when assessing surgical technical skill using VR simulators.

The essential measures of a psychometrically sound test are its reliability and its validity. The validity of a test must be considered proportional to the realism of the simulation. Few studies have focused on the validation of VR training tools. So far, only a few research groups have attempted to validate a VR simulator in a setting that goes beyond the context of a *basic* psycho-
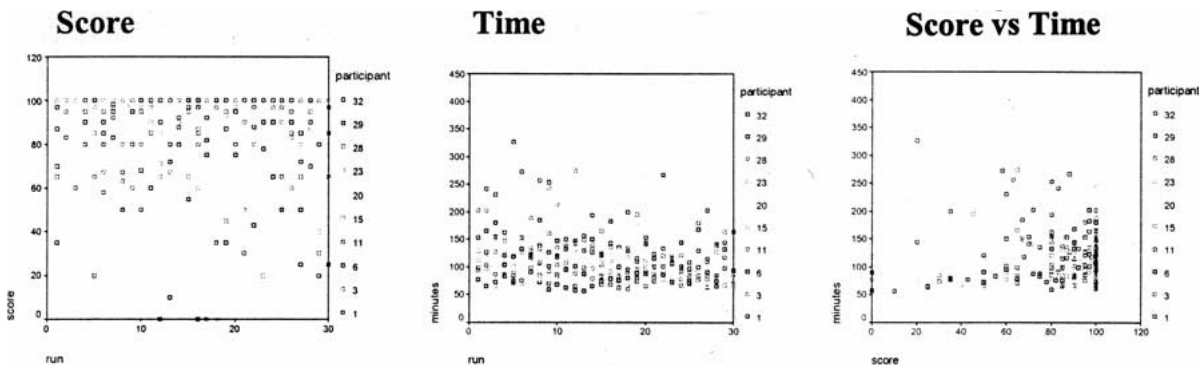
Fig. 9. Group 3. Moderate level of innate abilities, gaining unstable improvement through virtual reality training.
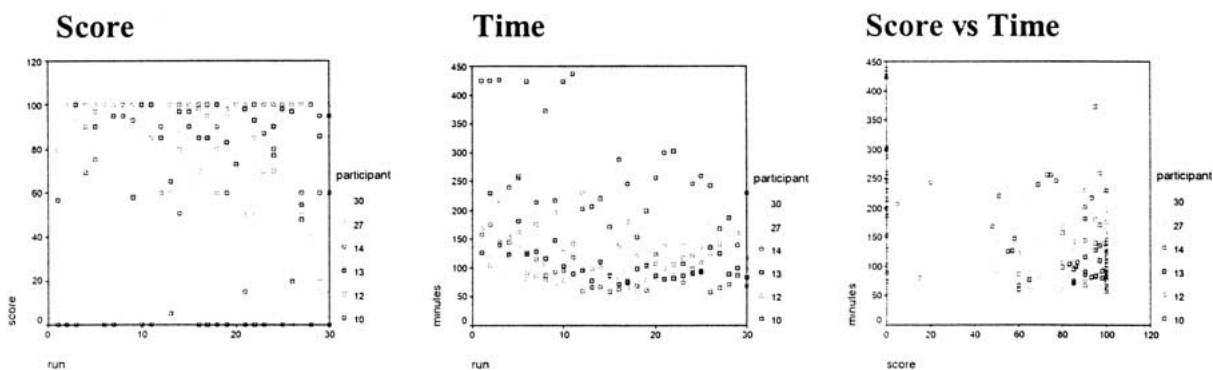

Fig. 10. Group 4. Low level of innate abilities, not gaining improvement through virtual reality training.

motor skills trainer. No studies have attempted to assess the learning curve for laparoscopic *task-oriented procedural* VR settings. It must be stressed that the term "learning curve" is, in fact, a misnomer. Learning is a parameter that cannot be measured in itself. It is usually an extrapolation from changes in performance over time. It is important to realize that the outcome measures used in this study—performance time and sumscore—are fixed due to time restraints and the finite endpoint of score.

There are four major implications of our study that result from the assessment of the learning curves for our VR procedural surgical simulation. First, the data show that is not opportune to assess a VR system's learning curve using only one test subject who performs a simulation task repeatedly. In fact, there appear to be very different profiles in the spectrum of performers. We identified four basic profiles of performers along the spectrum. Performers in groups 1 and 4 have profiles that do not seem to improve through repetitive training. The explanation for this phenomenon is, in essence, very different for the two groups. Group 1 appears to have such strong innate abilities that after the 1-h hands-on familiarization protocol, there is little more to be gained from repetitive training. This group comprises 16.7% of the total group. In contrast, in group 4 (20% of the total group), native abilities seem to be lacking, so there is not much of an innate psychomotor framework to build upon.

Second, the study shows that most performers are indeed responsive to training. These are the individuals in groups 2 and 3 (together, 63.3% of the total group). Performers in these groups displayed curves with a definite asymptote, although performance is less stable in group 3 than in group 2. An equilibrium occurs after ~25 runs, after which no major errors are incurred. From the plots, we can conclude that the parameter of time, in itself, is a fairly untrustworthy basis for assessing surgical competence.

Third, the study shows that it is, in fact, possible to identify different profiles of performers using VR procedural simulation. This finding justifies the use of VR procedural simulators such as the Xitact LS 500 to train most surgical residents in the performance of the tasks essential for MAS procedures, such as the laparoscopic cholecystectomy.

Finally, it suggests that VR procedural simulators can be used for the future guidance and selection of surgical trainees in MAS.

The dynamics of acquiring laparoscopic proficiency through VR simulation are complex. In MAS, learning is likely to be affected by a variety of factors, including previous experience with the specific procedure or similar procedure, the experience of the supporting surgical team, the type of equipment, and of course the nature of the clinical case itself. Many of these factors are not of influence whilst training with the Xitact LS 500. None of the participants in this study had any hands-on experience in MAS. No anatomical variations were encountered in the repetitive simulations, in sharp contrast to clinical practice with real patients. The instruments did not vary during the simulation, nor did the place of

"operation." Excluding these factors, there is little statistical noise—e.g., validation bias—to be expected in our study. Nevertheless, these factors are definitively present in real-time operations.

We believe that it is important to proceed carefully, taking a stepwise approach in assessing VR simulative training through validation studies. Only by repeated validation in different settings, a solid and optimal framework for the use of VR simulation be built. Therefore, further research should be aimed at the elaboration of heretofore undefined, and therefore critical parameters, parameters such as the ideal time interval for initial training in VR simulators during the learning curve and the establishment of optimal training schedules thereafter.

The ultimate purpose of VR simulations of laparoscopic cholecystectomy is not to train residents to operate safely on simulators; it is to train residents to operate safely on patients. Ultimately, the endpoint of a valid and stepwise VR validation, and of learning skill studies in general, is to address the question of whether the skill acquired on the simulator in fact translates to the clinical setting.

Advances in medical science and technology are likely to be accompanied by dramatic changes in the way that surgery is taught. For many procedures, the surgical community is now moving from the open surgical approach to the minimally invasive one, where different sensory feedback properties are eminent and specific psychomotor skills are important. Such a paradigm shift will have profound implication for the way surgical training programs are developed, surgeons are selected, and policy for the (re)certification surgeons is established.

References

1. Ali MR, Mowery Y, Kaplan B, De Maria EJ (2002) Training the novice in laparoscopy: more challenge is better. Surg Endosc 16: 1723–1736
2. Bevan PG (1986) Crafts workshops in surgery. Br J Surg 73: 1–2
3. Coleman J, NDuka CC, Darzi A (1994) Virtual reality and laparoscopic surgery. Br J Surg 81: 1709–1711
4. Cosman PH, Cregan PC, Martin CJ, Cartmill JA (2002) Virtual reality simulators: current status in acquisition and of surgical skills. Aust AN Z J Surg 72: 30–34
5. Deziel D, Milikan KW, Economon SG, Doolas A, Ko ST, Airan MC (1993) Complications of laparoscopic cholecystectomy: a national survey of 4,292 hospitals and an analysis of 77,604 cases. Am J Surg 165: 9–14
6. Farrell A (2002) Can virtual reality be used to measure and train surgical skills? Ergonomics 45: 362–379
7. Gallagher A, Satava M (2002) Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Surg Endosc 16: 1746–1752
8. Grantcharov TP (2002) Teaching and testing surgical skills on a VR laparoscopy simulator: learning curves and effect of previous operative experience on performance. 10th International Congress of the European Association for Endoscopic Surgery (EAES), Lisbon, Portugal
9. Grantcharov TP, Baruran L, Fuch-Jensen P, Rosenberg J (2003) Learning curves and impact of previous operative experience on performing on a virtual reality simulator to test laparoscopic surgical skills. Am J Surg 185: 146–149
10. Hamdorf JM, Hall JC (2000) Acquiring surgical skills. Br J Surg 87: 28–37
11. Melvin SW, Johnson JA, Ellison CE (1996) Laparoscopic skills enhancement. Am J Surg 172: 377–379
12. Moore MJ, Bennett CL (1995) The learning curve for laparoscopic cholecystectomy. Am J Surg 170: 55–59
13. Peters JH, Ellison EC, Innes JT (1991) Safety and efficacy of laparoscopic cholecystectomy. Ann Surg 312: 3–12
14. Ramsay CR, et al. (2001) Statistical assessment of the learning curves of health technologies. Health Technol Assess 5: 1–98
15. Satava MR, Cushieri A, Hamdorf J (2003) Metrics for objective assessment: preliminary summary of the Surgical Skills Workshop. Surg Endosc 17: 220–2226
16. Schijven M, Jakimowicz J (2002) Face, expert, and referent validity of the Xitact LS 500 laparoscopy simulator. Surg Endosc 16: 1764–1770
17. Schijven M, Jakimowicz J (2002) Construct validity: experts and residents performing on the Xitact LS 500 laparoscopy simulator. Surg Endosc (In press)
18. Schijven MP, Jakimowicz J, Schot C (2002) The Advanced Dundee Endoscopic Psychomotor Tester (ADEPT) objectifying subjective psychomotor test performance. Surg Endosc 16: 943–948
19. Southern Surgeons Club (1995) The learning curve for laparoscopic cholecystectomy. Am J Surg xx: 55–59
20. Strorn P, et al. (2003) Validation and learning in the Procedicus KSA virtual reality surgical simulator: implementing a new safety culture in medical school. Surg Endosc 17: 227–231
21. Wanzel KR, Ward M, Reznick RK (2002) Teaching the surgical craft: from selection to certification. Curr Probl Surg 39: 574–659