Can skills assessment on a virtual reality trainer predict a surgical trainee's talent in laparoscopic surgery?

R. Rosenthal,[1] W. A. Gantert,[1] D. Scheidegger,[2] D. Oertli[1]

[1] Department of Surgery, Universitätsspital Basel, 4031 Basel, Switzerland
[2] Department of Anesthesia, Universitätsspital Basel, 4031 Basel, Switzerland

Abstract

Background: A number of studies have investigated several aspects of feasibility and validity of performance assessments with virtual reality surgical simulators. However, the validity of performance assessments is limited by the reliability of such measurements, and some issues of reliability still need to be addressed. This study aimed to evaluate the hypothesis that test subjects show logarithmic performance curves on repetitive trials for a component task of laparoscopic cholecystectomy on a virtual reality simulator, and that interindividual differences in performance after considerable training are significant. According to kinesiologic theory, logarithmic performance curves are expected and an individual's learning capacity for a specific task can be extrapolated, allowing quantification of a person's innate ability to develop task-specific skills.
Methods: In this study, 20 medical students at the University of Basel Medical School performed five trials of a standardized task on the LS 500 virtual reality simulator for laparoscopic surgery. Task completion time, number of errors, economy of instrument movements, and maximum speed of instrument movements were measured.
Results: The hypothesis was confirmed by the fact that the performance curves for some of the simulator measurements were very close to logarithmic curves, and there were significant interindividual differences in performance at the end of the repetitive trials.
Conclusions: Assessment of perceptual motor skills and the innate ability of an individual with no prior experience in laparoscopic surgery to develop such skills using the LS 500 VR surgical simulator is feasible and reliable.

Key words: Laparoscopy — Simulation — Training — Virtual reality

---

Correspondence to: R. Rosenthal

Traditionally, surgical skills training has been based on the apprenticeship model introduced by William Halsted more than 90 years ago. The surgical trainee learned surgical skills primarily in the operating room by observing, assisting, and taking a more and more active role in the procedure as his or her experience grew. However, the recent introduction of high-technology surgery (e.g., endoscopic/laparoscopic and robotic surgery), which requires additional surgical skills, has exposed the significant shortcomings of traditional apprenticeship training in the operating room [1–3].

Ethical, economic, and educational considerations have led to the development of alternative methods for teaching and training surgical technique [4, 5]. We can deduct from hierarchical task decomposition and analyses of surgical performance using an information processing model that certain basic perceptual motor skills, defined by information acquisition and motor output, and a number of higher cognitive skills are important in completing laparoscopic surgical tasks [6, 7].

In recent years, several virtual reality simulators have been developed that simulate either simple abstract manipulations typically used in endoscopic surgery or highly realistic component tasks of laparoscopic operations with real-time graphic, with some simulators providing haptic representation [8–10]. To date, the main focus of simulator development has been on training and objectively assessing basic perceptual motor skills in endoscopic surgery [11–13].

A number of research groups worldwide have investigated several aspects of feasibility and validity of performance assessments with virtual reality surgical simulators [14–19]. However, the validity of performance assessments is limited by the reliability of such measurements, and some issues related to the reliability of simulator performance assessments still need to be addressed [20]. One important question is how repetitive performance of a task on the surgical simulator influences performance assessments (i.e., what impact does learning have on performance assessment). Another

important issue in performance assessment is whether simulator measurements are sufficiently sensitive to detect interindividual differences in innate ability to develop adequate skills.

Kinesiologic research has detected a number of important principles underlying motor performance and learning, improving our understanding of the relationship among talent, acquired skills, and the learning processes underlying skill acquisition [21]. Abilities are defined as genetically determined and unmodifiable by practice or experience [21]. They are of limited number, stable, and enduring, and they underlie many different skills. Examples of abilities include multilimb coordination, finger dexterity, arm–hand steadiness, kinesthetic sensitivity, spatial orientation, and reaction time.

Skills, on the other hand, describe an individual's proficiency at a particular task, are developed and modified with practice, are countless in number, and depend on several abilities [21]. One important finding is that improvement in motor performance over repetitive trials of a specific task is highly predictable and follows a logarithmic function. An individual's learning capacity for a specific task can therefore be extrapolated by measuring performance during only a few repetitive trials. This allows quantification of an individual's talent or, in kinesiologic terms, innate ability to develop task-specific skills. In addition, kinesiologic research has demonstrated that individuals differ significantly in their perceptual motor abilities, with individual differences defined as stable differences among individuals on some variable or task [22].

We designed a study to investigate some aspects of the reliability of performance measurements provided by the Xitact LS 500 Virtual Patient, a simulator designed to train laparoscopic procedures. Our study aimed to investigate whether the perceptual motor skills and the ability of test subjects without prior experience in endoscopic surgery could be reliably assessed on the basis of performance measurements provided by the LS 500 simulator. We hypothesized that performance measurements of perceptual motor skills over repetitive trials of a task on the LS 500 simulator would follow the principles of kinesiologic theory, and that these measurements would be sufficiently sensitive to detect interindividual differences in the perceptual motor abilities of test subjects.

Methods

All the experiments in the current study were conducted at the Swiss Center for Medical Simulation, a designated training facility at the University Hospital of Basel, integrating an anesthesia simulator with the Xitact LS 500 surgical simulator.

The LS 500 Virtual Patient laparoscopy simulator consists of a Pentium PC with a 19-in. high-resolution TFT monitor connected to two robotic force feedback devices that act as bidirectional interfaces for the laparoscopic instruments, and to a third unidirectional electromechanical interface that directs the laparoscope. The proprietary software runs on a Windows 2000 operating system and simulates three steps of a laparoscopic cholecystectomy with real-time, high-resolution graphic and haptic representation of the operative field (exposure, dissection, and clip-and-cut cystic duct/artery). A choice of laparoscopic instruments is simulated including graspers, scissors, clip applier, and hook electrode with realistic instrument handles connected to the manipulation robots. The software records global task completion time, completion time of subtasks, maximum speed of the dominant and nondominant hand instrument, tool tip travel distance of both instruments, errors in clip placement, and errors in cutting.

Test subjects

For the study, 20 medical students at the University of Basel Medical School with no prior experience in laparoscopic surgery were recruited at random. All the participants completed a standardized questionnaire of items asking them about their career goal and their prior experience with video games, then gave informed consent to participate in the study. In addition, their stereopsis was examined using the Lang Stereotest I. To validate the possibility of extrapolating performance, four residents performed the same simulator task 18 times.

Task

Because the component task of clipping and cutting the cystic duct and artery is a crucial task of laparoscopic cholecystectomy, it was chosen as the test item. Each participant performed a series of five clip-and-cut tests on the Xitact LS 500 VR simulator.

Statistical analysis

The data were analyzed using SPSS 8.0 statistical software. Logarithmic regression analyses of the repetitive trial data were performed to compare the data with the expected logarithmic performance curves. The Kruskal–Wallis test was used to test interindividual differences. Differences between groups were evaluated using the chi-square test for ordinal data and the two-tailed t-test for continuous data. Statistical significance was defined as a p value less than 0.05.

Results

The study participants were 8 women and 12 men with an average age of 26 years. A total of 8 test subjects (40%) indicated experience with videogames: 2 (25%) of the 8 female and 6 (50%) of the 12 male participants. All 20 students had optimal stereoptic capability. Whereas 5 of the participants mentioned surgery as their intended speciality, the remaining 15 mentioned other medical specialties. The participants included 4 residents (2 women and 2 men) with an average age of 30 years and an average of 3.8 years experience in surgery. All the test subjects finished all the trials of the clip-and-cut task.

Performance curves

Students

All the performance measurements were plotted against the trial number. A logarithmic performance curve was found with the following performance measurements: time (time to first and last clip or cut, total task time), economy of movement (left and right instrument total distance), and performance in clipping (clips used, clips lost, and clipping of bad organ). Figure 1 shows an example of the "time to first cut" performance measurement plotted against the trial number. Table 1 gives an overview of the regression coefficients of the logarithmic regression analyses.
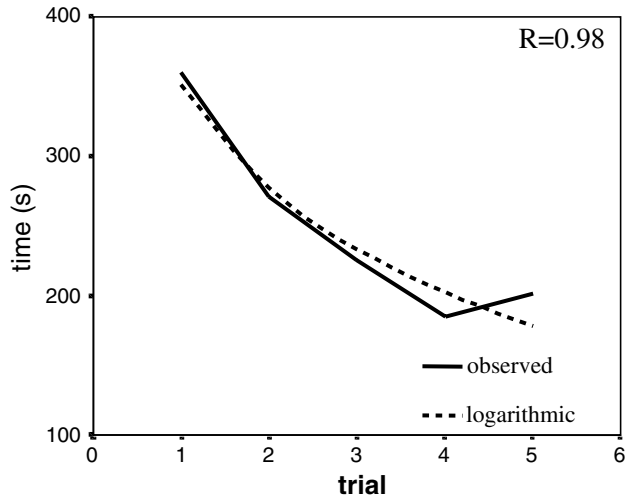
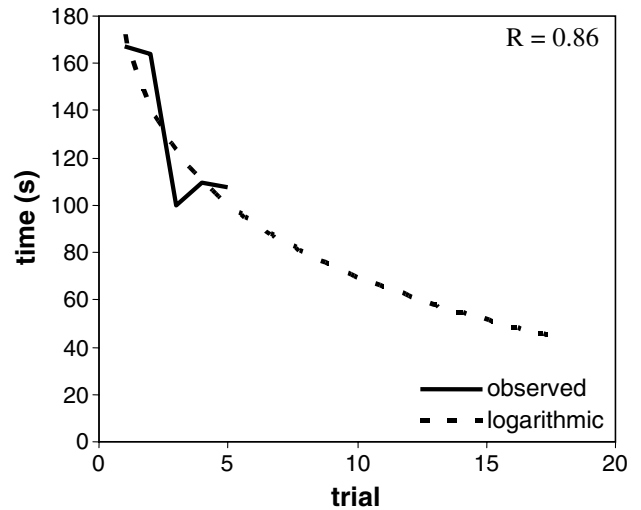**Fig. 1.** Time to first cut plotted against trial number (mean of all students).

Table 1. Regression coefficients of the performance measurements

	r
Time to first clip	0.92
Time to last clip	0.97
Time to first cut	0.98
Time to last cut	0.98
Total task time	0.88
Left instrument total distance	0.84
Right instrument total distance	0.75
Left instrument maximum speed	0.10
Right instrument maximum speed	0.31
Clips lost	0.88
Clipping of bad organ	0.90
Cutting of bad organ	0.44
Clips used	0.86

Residents

The calculated logarithmic trend line for the parameter task "completion time" after five trials ($R = 0.86$) (Fig. 2) was almost identical to the one found after 18 trials ($R = 0.89$) (Fig. 3). As shown in the figure, the flat part of the curve was reached after 18 trials.

Interindividual differences

Performance measurements showed significant interindividual differences at the end (last three trials) of each series of repetitive trials with regard to the economy of movement of both instruments and the time for the task. Table 2 depicts the p values for the Kruskal–Wallis tests.

Differences between groups

Performance was compared between female and male test subjects. The results showed no statistical significance, but a tendency for better performance was observed in the male group. The comparison of the subjects who had experience in videogames with those
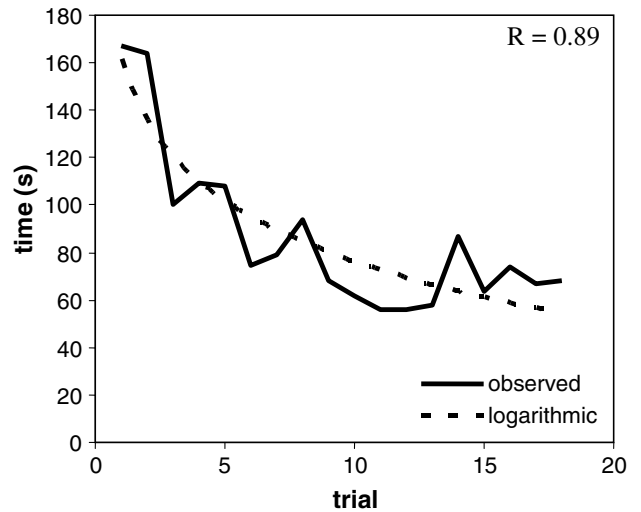

Fig. 2. Task completion time plotted against trial number, 5 trials (mean of all residents).

Fig. 3. Task completion time plotted against trial number, 18 trials (mean of all residents).

who had no such experience showed a tendency toward better performance in the group with experience, but the results were not statistically significant.

Discussion

Our study shows that the highly realistic clip-and-cut task of the Xitact LS 500 surgical simulator can be performed by medical students without any prior experience in laparoscopic surgery (feasibility); that the performance curves for some of the performance measurements recorded during a series of repetitive trials represent almost perfect logarithmic performance curves, as expected by kinesiologic theory (reliability); and that these performance measurements are sufficiently sensitive to detect interindividual differences among test subjects even after considerable practice (a difference in innate ability to develop perceptual motor skills apparently is important in laparoscopic surgery).

Table 2. The p values for interindividual differences

	p Value for last three trials
Time to first clip	NS
Time to last clip	0.011
Time to first cut	0.046
Time to last cut	0.013
Total task time	NS
Left instrument total distance	0.025
Right instrument total distance	0.002
Left instrument maximum speed	NS
Right instrument maximum speed	NS
Clips lost	NS
Clipping of bad organ	NS
Cutting of bad organ	NS
Clips used	NS

NS, not significant

In the recent past, several research groups have shown that simple abstract tasks on a virtual reality trainer can be performed by inexperienced surgical novices [8, 11]. In one study, inexperienced surgical trainees also were able to perform the clip-and-cut task on the LS 500 [16]. We chose medical students as test subjects because career planning for a medical specialty usually is taking place at the medical student level. Accordingly, a fair selection process aimed at selecting talented surgical trainees should be adapted to the medical student level.

Our study has clearly shown that highly realistic tasks on a virtual reality simulator can be used to select medical students with appropriate talent for developing the skills necessary in endoscopic surgery. We chose the clip-and-cut task for the following reasons. First, it is a simple task that may be performed by students with no prior experience in surgery. Second, by having the students perform a circumscribed task, the comparison between the students was more reliable. Current software packages provide programs for whole procedures such as cholecystectomy or treatment of an ectopic pregnancy and more. These may be used for the assessment and training of more advanced surgeons.

Learning is not directly observable. Only motor performance can be quantified in terms of speed and precision. Learning is therefore represented indirectly by improvement in motor performance at a certain task with repetitive trials [21].

The fact that the performance curves for some of the performance measurements, when plotted in comparison with trial number, follow a logarithmic curve as predicted by kinesiologic theory is remarkable. It enables us to predict the trainee's learning potential for a component laparoscopic skill. By measuring performance during only a few repetitive trials of the task, we can predict the trainee's task performance score after an extended period of repetition and training. Although the innate abilities underlying performance in laparoscopic surgery still are largely unknown, we at least have a way to estimate an individual's ability to develop the specific perceptual motor skills necessary for laparoscopic task completion.

At first glance, our findings appear in sharp contrast to the results of another similar study reported by Schijven and Jakimowicz [16]. However, as pointed out previously, only some of the performance measurements in our study followed a logarithmic function. The simulator measurements that did not show logarithmic improvement over repetitive trials appear to be unreliable. For example, we did not find a logarithmic performance curve with regard to the maximum speed of the left- and right-hand instruments. This indicates that the maximum speed of an instrument is not a reliable and valid parameter of adequate performance. The movement of an instrument in expert performance may be very precise and rather slow, as compared with a very fast instrument move by a novice that may cause more harm than good.

There are several other possible reasons why some performance curves do not follow a logarithmic function. A ceiling or floor effect of performance assessment may prevent further improvement because the maximal possible score has already been reached. There may be too much random variation from trial to trial, or the measurement techniques may be inappropriate (e.g., ordinal data measurement), that is, not sensitive enough to detect performance improvement.

Such limitations of performance assessments are well known from other tests of psychomotor performance [20, 23]. An analysis of the results reported by Schijven et al. shows that they actually confirm the results of our study because these authors also found that some performance measurements followed predictive patterns and some did not [16].

For decades, researchers have tried to develop methods for objective assessment of surgical technical skills and innate ability. Reznick [24] and his colleagues proposed five ways that technical skills can be quantified: procedure listing with logs, direct observation, direct observation with criteria, videotaping, or use of an animal or simulation model. Despite their high validity, the first four methods are unlikely to be used in clinical routine because of their complexity and their impracticability for repetitive assessment. In this respect, the virtual reality simulator has a clear advantage with its permanent availability and the possibility for unrestricted task repetitions. Our study has shown that performance assessment on a fairly small number of repetitive trials of a component task is sufficient to estimate an individual's ability to develop adequate skills.

Comparing the group that had experience in videogames with the group that had no such experience, we did not find a significant difference, only a tendency to better performance in the experienced group. The same must be said concerning the difference between male and female participants. The male group contained more test subjects with experience in videogames, which may have contributed to the tendency for better performance. Yet the number of participants was too small for a definitive statement to be made.

Our study had a number of drawbacks. The number of repetitive trials that each test person performed was relatively small. However, because the performance curves for some performance measurements showed highly predictable patterns, it is very unlikely that further repetitions of task performance would have yielded a

different result. The logarithmic trend line of the four residents repeating the task 18 times is almost identical to the trend line for 5 trials, confirming this assumption.

Further studies are needed to investigate the nature and impact of cognitive skills involved in endoscopic surgical task performance, such as spatial abilities, memory, knowledge, attention, and reasoning, and to evaluate the feasibility, validity, and reliability of proposed performance assessments for such higher cognitive skills on virtual reality simulators. Some of these issues are currently under investigation by our group.

Conclusion

We conclude that use of the LS 500 VR surgical simulator to assess and train perceptual motor skills and to assess innate ability for the development of such skills by a subject with no prior experience in laparoscopic surgery is feasible and reliable. Interindividual differences in ability can be assessed with some performance measurements over only a few repetitive trials.

References

1. Haluck RS, Krummel TM (2000) Computers and virtual reality for surgical education in the 21st century. Arch Surg 135: 786–792
2. Marescaux J, Leroy J, Gagner M, Rubino F, Mutter D, Vix M, Butner SE, Smith MK (2001) Transatlantic robot-assisted telesurgery. Nature 413: 379–380
3. 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
4. Gorman PJ, Meier AH, Krummel TM (1999) Simulation and virtual reality in surgical education: real or unreal? Arch Surg 134: 1203–1208
5. Bridges M, Diamond DL (1999) The financial impact of teaching surgical residents in the operating room. Am J Surg 177: 28–32
6. Tendick F, Downes M, Cavusoglu M, Gantert W, Way LW (1998) Development of virtual environments for training skills and reducing errors in laparoscopic surgery. SPIE Conference 3262 on Surgical-Assist Systems. San Jose, CA, pp 36–44
7. Cao C, MacKenzie CL, Ibbotson JA, Turner LJ, Blair NP, Nagy AG (1999) Hierarchical decomposition of laparoscopic procedures. Stud Health Technol Inform 62: 83–89
8. Gallagher AG, McClure N, McGuigan J, Crothers I, Browning J (1999) Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical traininer virtual reality (MIST-VR). Endoscopy 31: 310–313
9. Szekely G, Satava RM (1999) Virtual reality in medicine. BMJ 319: 1305
10. O'Toole RV, Playter RR, Krummel TM, Blank WC, Cornelius NH, Roberts WR, Bell WJ, Raibert M (1999) Measuring and developing suturing technique with a virtual reality surgical simulator. J Am Coll Surg 189: 114–127
11. Chaudhry A, Sutton C, Wood J, Stone R, McCloy R (1999) Learning rate for laparoscopic skills on MIST-VR, a virtual simulator: quality of human–computer interface. Ann R Coll Surg Engl 81: 281–286
12. Taffinder NJ, McManus IC, Gul Y, Russell RCG, Darzi A (1998) Effect of sleep deprivation on surgeons' dexterity on laparoscopy simulator. Lancet 352: 1191
13. Darzi A, Smith S, Taffinder N (1999) Assessing operative skill: needs to become more objective. BMJ 318: 887–888
14. Devitt JH, Kurrek MM, Cohen MM, Cleave-Hogg D (2001) The validity of performance assessments using simulation. Anesthesiology 95: 36–42
15. Jordan JA, Gallagher AG, McGuigan J, McGlade K, McClure N (2000) A comparison between randomly alternative imaging, normal laparoscopic imaging, and virtual reality training in laparoscopic psychomotor skill acquisition. Am J Surg 180: 208–211
16. Schijven MP, Jakimowicz J (2004) The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. Surg Endosc 18: 121–127
17. Schijven MP, Jakimowicz J (2002) Face, expert, and a referent validity of the Xitact LS 500 laparoscopy simulator. Surg Endosc 16: 1764–1770
18. Schijven MP, Jakimowicz J (2003) Construct validity: experts and residents performing on the Xitact LS500 laparoscopy simulator. Surg Endosc 17: 803–810
19. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P (2004) Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 91: 146–150
20. Risucci DA, Tortolani AJ (1990) A methodological framework for the design of research on the evaluation of residents. Acad Med 65: 36–41
21. Schmidt RA (1982) Motor control and learning: a behavioral emphasis. Human Kinetics Publishers, Champaign, IL
22. Ivry R (1996) Representational issues in motor learning: phenomena and theory. In: Handbook of perception and action. Academic Press, San Diego, CA
23. Hanna GB, Drew T, Clinch P, Shimi S, Dunkley P, Hau C, Cuschieri A (1997) Psychomotor skills for endoscopic manipulations: differing abilities between right- and left-handed individuals. Ann Surg 225: 333–338
24. Reznick R (1993) Teaching and testing technical skills. Am J Surg 165: 358–361