Positive Correlation Between Motion Analysis Data on the LapMentor Virtual Reality Laparoscopic Surgical Simulator and the Results from Videotape Assessment of Real Laparoscopic Surgeries

Tadashi Matsuda, M.D.,¹ Elspeth M. McDougall, M.D., FRCSC, MHPE,² Yoshinari Ono, M.D.,³ Ryohei Hattori, M.D.,⁴ Shiro Baba, M.D.,⁵ Masatsugu Iwamura, M.D.,⁵ Toshiro Terachi, M.D.,⁶ Seiji Naito, M.D.,⁷ and Ralph V. Clayman, M.D., FACS²

Abstract

Purpose: We studied the construct validity of the LapMentor, a virtual reality laparoscopic surgical simulator, and the correlation between the data collected on the LapMentor and the results of video assessment of real laparoscopic surgeries.

Materials and Methods: Ninety-two urologists were tested on basic skill tasks No. 3 (SK3) to No. 8 (SK8) on the LapMentor. They were divided into three groups: Group A (n=25) had no experience with laparoscopic surgeries as a chief surgeon; group B (n=33) had <35 experiences; and group C (n=34) had ≥35 experiences. Group scores on the accuracy, efficacy, and time of the tasks were compared. Forty physicians with ≥20 experiences supplied unedited videotapes showing a laparoscopic nephrectomy or an adrenalectomy in its entirety, and the videos were assessed in a blinded fashion by expert referees. Correlations between the videotape score (VS) and the performances on the LapMentor were analyzed.

Results: Group C showed significantly better outcomes than group A in the accuracy (SK5) ($P=0.013$), efficacy (SK8) ($P=0.014$), or speed (SKs 3 and 8) ($P=0.009$ and $P=0.002$, respectively) of the performances of LapMentor. Group B showed significantly better outcomes than group A in the speed and efficacy of the performances in SK8 ($P=0.011$ and $P=0.029$, respectively). Analyses of motion analysis data of LapMentor demonstrated that smooth and ideal movement of instruments is more important than speed of the movement of instruments to achieve accurate performances in each task. Multiple linear regression analysis indicated that the average score of the accuracy in SK4, 5, and 8 had significant positive correlation with VS ($P=0.01$).

Conclusions: This study demonstrated the construct and predictive validity of the LapMentor basic skill tasks, supporting their possible usefulness for the preclinical evaluation of laparoscopic skills.

Introduction

SINCE THE DEVELOPMENT of laparoscopic cholecystectomy, nephrectomy,¹ and adrenalectomy,² laparoscopic surgeries have become the mainstream of surgery, promoting minimally invasive therapies. Laparoscopic surgeries need different technical demands from the conventional open surgeries.

To prevent complications, it is mandatory to develop a system to assess the competency of each laparoscopic surgeon; however, it is a difficult task to assess the competency of surgeons, and to date, no reliable system has been created.³ Intheater assessment by supervisors has been performed for more than 200 years, but in newly developed techniques such as laparoscopy, the competency of the supervisors also needs to be evaluated.

¹Department of Urology and Andrology, Kansai Medical University, Hirakata, Japan.

²Department of Urology, University of California Irvine Medical Center, Orange, California.

³School of Medical Welfare, Aichi Shukutoku University, Nagoya, Japan.

⁴Department of Urology, Nagoya University School of Medicine, Nagoya, Japan.

⁵Department of Urology, Kitasato University School of Medicine, Sagami, Japan.

⁶Department of Urology, Tokai University School of Medicine, Isezaki, Japan.

⁷Department of Urology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan.

In an effort to resolve these problems, the Japanese Society of Endoscopic Surgery developed a unique system to assess the ability of laparoscopic surgeons: The Endoscopic Surgical Skill Qualification (ESSQ) system.⁴ Double-blinded, laparoscopically experienced referees assess the ability of a surgeon by viewing videotapes showing a complete, unedited laparoscopic nephrectomy or laparoscopic adrenalectomy by the applicant. Two referees were randomly selected for each video among a pool of 29 referees, each of whom had performed more than 100 urologic laparoscopic surgeries. Video assessment is based on procedural safety (ie, complication-free maneuvers) rather than on the speed of the procedure. When there are errors or dangerous maneuvers in the laparoscopic procedures, points are deducted. Each surgeon begins with 75 points; if more than 15 points are deducted, the surgeon fails the assessment.

Since the start of the ESSQ system in 2004, 1308 Japanese urologic laparoscopists have been evaluated; the pass rate is 60.2% (unpublished data). This method of evaluation, however, is extremely time consuming for the referees; on average, 3 hours were spent evaluating each of the 1308 applicants.

Accordingly, what is needed is a simpler system to assess technical proficiency, thereby ensuring patient safety. To this end, virtual reality (VR) simulators could be quite useful, especially if they could be correlated with actual surgical performance.⁵ A variety of VR simulators have been developed in the field of laparoscopy. LapMentor™ (Simbionix Ltd, Lod, Israel) is a computer-based VR simulator, featuring two mock working instruments and a simulated laparoscopic endoscope.

To determine the true usefulness of VR simulators in surgical education, validity studies are mandatory, especially of a construct and predictive nature. Confirmation of construct validity would demonstrate that the score on the VR simulator performance could distinguish the surgeons based on their experience with the surgical technique. The ultimate test of a simulator is predictive validity, which demonstrates that based on the surgeon's performance score on the simulator, the clinical proficiency could be reliably predicted for that surgeon in performing the operative procedure.⁶

We performed a study to evaluate both construct and predictive validity of LapMentor basic tasks among 92 Japanese urologists. The correlation between the motion analysis data on LapMentor for several skills and the results of video assessment in the ESSQ system (construct and predictive validity) as well as the experience of the surgeon (construct validity) were assessed.

Materials and Methods

Ninety-two Japanese urologists were enrolled in the study. The average age of the subjects was 38.5 years, ranging from 23 to 58 years. The median number of laparoscopic surgeries performed by these physicians as a chief surgeon was 25, with a range of 0 to 700 procedures. Subjects were divided into three groups according to their laparoscopic experience: Group A consisted of 25 physicians with no experience as chief surgeons in laparoscopic surgeries. The rest of the subjects were divided into two groups, each of which had almost a similar number of subjects. Group B consisted of 33 physicians who had performed 1 to 34 laparoscopic surgeries as

chief surgeons, and group C consisted of 34 physicians who had performed 35 or more surgeries (Table 1).

Forty physicians in our study, each with >20 laparoscopic surgeries accepted to supply an unedited videotape showing themselves performing either a laparoscopic nephrectomy or a laparoscopic adrenalectomy in its entirety. Twelve of them belonged to group B and 28 belonged to group C. Videos of 30 nephrectomies and 10 adrenalectomies were supplied. Each video was assessed blindly (without information on the operator's name and the results of LapMentor tasks by the operator) by one referee randomly selected out of six expert referees according to the ESSQ system evaluation criteria.⁵ Perfect procedures with no dangerous maneuvers in the entire operative procedures achieved a perfect videotape score (VS) of 75 points. The six referees had been working as referees of the ESSQ system for 3 years after more than two consensus meetings on video assessment guidelines. After initiation of the system, the referees attended at least 1 consensus meeting every year. At the consensus meetings, videos showing inappropriate procedures are reviewed and consensus is established on how many points should be deducted for each inappropriate procedure.

Basic skill tasks No. 3 to No. 8 (SK 3 to 8) on the LapMentor were performed by all 92 surgeons without any pretest exercises. None of the subjects had performed LapMentor basic tasks before the test. SK3 requires an examinee to touch a round ball by the tip of right or left hand forceps according to the color of the balls (Fig. 1A). SK4 requires an examinee to place a clip on a tube within a specific highlighted segment. SK5 is the same setting as SK4, except that the examinee clips a tube while simultaneously pulling the tube with a forceps in the other hand (Fig. 1B). SK6 requires finding a ball in a gel-like mountain, grasping the ball, and transferring it to a basket. SK7 requires the examinee to cut a round shape from a virtual sheet of material using a straight scissors. SK8 requires an examinee to coagulate a single highlighted band among 12 bands using a right-handed or left-handed hook electrode that is keyed with a respective right or left foot pedal. After each band is coagulated, the examinee is provided with a new highlighted band until all 12 bands have been randomly highlighted and been electrocoagulated. Errors occur if the wrong band is coagulated with either of the hook electrodes (Fig. 1C).

Data including times for each performance, percentages of errors, times of energy delivery, speed of movements of instruments, number of motions of instruments, efficacy of movements of instruments shown as ratios of the length of ideal movements calculated by LapMentor and that of real movements, and percentage of success in the required performances were recorded by the LapMentor. The following parameters in each task were used as the representative ones of the quality of the performance: The average time needed to touch one ball (Time SK3) and the accuracy in touching the ball in SK3 (Accuracy SK3), the total time needed to complete the required task (Time SK4) and the accuracy of the clip application in SK4 (Accuracy SK4), the total time to complete the task (Time SK5) and the accuracy of the clip application in SK5 (Accuracy SK5), the average time to move one ball in SK6 (Time SK6), the total time to complete the task (Time SK7) and the accuracy in cutting the sheet without damage in SK7 (Accuracy SK7), and the average time to coagulate one band (Time SK8), the efficacy of the electrocoagulation (Efficacy

TABLE 1. CHARACTERISTICS AND LAPMENTOR MOTION ANALYSIS DATA OF THE THREE STUDY GROUPS

Group	A	B	C	ANOVA	Tukey multicomparison post hoc test		
No. of participants	25	33	34		GA vs GB	GA vs GC	GB vs GC
Age (range)	30.5±4.1 (23–39)	37.5±4.9 (28–48)	45.3±5.8 (36–58)	0.000	0.000	0.000	0.000
Year of laparoscopic experience (range)	0.9±1.3 (0–5)	4.0±2.2 (1–10)	7.4±3.6 (3–16)	0.000	0.000	0.000	0.000
No. of cases experienced as a chief surgeon (range)	0	20.5±9.7 (2–34)	142.2±154.5 (35–700)	0.000	0.000	0.000	0.000
SK3: Time needed to touch 1 ball (second)	5.8±1.3	5.2±1.2	4.8±1.2	0.013	0.205	0.009	0.355
SK3: Accuracy in touching the ball (%)	95.9±4.8	94.1±7.1	94.7±6.1	0.570			
SK4: Time needed to perform the required task (second)	66.8±11.7	67.2±10.4	74.1±16.1	0.048			
SK4: Accuracy in clipping the tube (%)	81.0±14.6	79.9±13.3	80.7±14.4	0.947			
SK5: Time needed to perform the required task (second)	121.5±34.2	106.7±19.9	107.9±25.4	0.077			
SK5: Accuracy in clipping the tube (%)	79.6±14.2	85.7±11.4	88.6±10.2	0.017	0.131	0.013	0.571
SK6: Time needed to move 1 ball (second)	16.0±4.7	15.5±3.9	14.4±5.5	0.407			
SK7: Time needed to perform the required task (second)	157.0±64.3	139.1±57.7	137.3±64.3	0.434			
SK7: Accuracy in cutting without damaging (%)	99.8±0.8	99.4±1.6	99.8±0.7	0.246			
SK8: Time needed to coagulate 1 band (second)	14.5±5.1	11.9±2.3	11.5±2.3	0.002	0.011	0.002	0.831
SK8: Efficiency in electrocoagulation (%)	85.2±7.1	88.9±4.8	89.2±4.2	0.010	0.029	0.014	0.958
SK8: Accuracy in coagulating the band (%)	98.8±2.1	98.7±2.7	99.0±2.8	0.884			

ANOVA = analysis of variance; GA = group A; GB = group B; GC = group C.

FIG. 1. LapMentor basic tasks. (A) SK3 requires an examinee to touch a round ball by the tip of right or left hand forceps according to the color of the balls. (B) SK5 requires an examinee to clip a tube within a specific segment by a clip with pulling the tube by another hand forceps. (C) SK8 requires an examinee to cut a color-changed string among 12 strings using a hook by energy delivery without touching the other strings.

SK8), and accuracy in coagulating the band in SK8 (Accuracy SK8). The correlation among these representative parameters and other motion analysis data in each task was analyzed.

Statistical analyses were performed on SPSS software. Comparisons between the three groups were performed by

one-way factorial analysis of variances test with Tukey multicomparison *post hoc* test. Correlation analyses were performed by Pearson tests. Multiple linear regression analysis with forward selection of variables was used to evaluate the effects of variables on VS. The Fisher exact probability test was used for 2×2 contingency table. A *P* value less than 0.05 was considered to be statistically significant.

Results

Comparison of performances on LapMentor between Groups A to C

Group C showed significantly better results than Group A in Time SK3, Accuracy SK5, Time SK8, and Efficacy SK8, while group B showed a better result than group A in Time SK8 and Efficacy SK8. There were no statistically significant differences between groups B and C (Table 1, Fig. 2).

Correlation among the representative parameters and other motion analysis data of LapMentor

The representative parameters of LapMentor tasks, which were indicated to be significant in the above analyses, Time SK3, Accuracy SK5, Time SK8, and Efficacy SK8, were analyzed. The results are shown in Table 2.

Correlation between video assessment points and LapMentor data

VS showed statistically significant positive or negative correlations with the clinical or the representative parameters in LapMentor tasks. Positive correlation was noted between Accuracy SK4, Accuracy SK5, Accuracy SK8, and VS ($R=0.333$, $P=0.036$, $R=0.320$, $P=0.044$, and $R=0.403$, $P=0.009$, respectively) and negative correlation between age, years of laparoscopic experience, Time SK4, and VS ($R=-0.350$, $P=0.027$, $R=-0.316$, $P=0.047$, and $R=-0.318$, $P=0.045$, respectively). The average of Accuracies SK4, SK5, and SK8 (Accuracy SK458) and VS showed positive correlation ($R=0.504$, $P=0.0009$) (Fig. 3). Multiple linear regression analysis evaluated among parameters including age, years of laparoscopic experience, number of cases experienced as a chief surgeon, and Accuracy SK458 showed that only Accu-

racy SK458 had positive correlation with VS ($P=0.01$) and the other parameters did not show significant correlations. VS was less than 60 points in 42.9% (3/7) when Accuracy SK458 was less than 85%, whereas if the Accuracy was $\geq 85\%$, only 3.0% (1/33) had VS less than 60 points ($P=0.013$, Fisher exact probability test).

Discussion

Our study demonstrated construct validity of the LapMentor VR simulator, distinguishing nonlaparoscopic physicians from novice and experienced laparoscopic surgeons. Speed, efficacy, or accuracy of performances in SKs 3, 5, and 8 of the LapMentor showed significant differences between the nonlaparoscopic physicians and the novice and experienced physicians.

McDougall and associates⁷ also demonstrated good construct validity of basic tasks of LapMentor. They showed that in SKs 3 to 6, the physicians showed better performances than the medical students, but there were no differences among residents, experienced physicians with low volume of surgeries (<30/year), and physicians with high-volume surgeries (>30). In SKs 7 and 8, experienced physicians with >30 laparoscopic surgeries/year showed significantly better results than physicians with <30 laparoscopic surgeries/year. McDougall and associates⁷ concluded that SK8 afforded the greatest distinction of surgical experience because of its degree of complexity.

In the present study, only Time SK8 and Efficacy SK8 showed significant difference between the novice physicians (group B) and the nonlaparoscopic physicians (group A), supporting the conclusion of McDougall and colleagues.⁷ We could not distinguish, however, the high-volume experienced physicians and the low-volume experienced physicians in any of the parameters of LapMentor. One reason for this discrepancy could be that in the study by McDougall and colleagues,⁷ the performances by physicians in a task were determined by the manufacturer's predetermined skill task scores, which were automatically calculated by the machine scoring software. In contrast, in our study, this built-in scoring calculation could not be used because of the difference of the software; thus, each set of motion analysis data was compared among the groups in our study. In addition, in our study, all

FIG. 2. Comparison of results of LapMentor basic tasks SK8 among Groups A to C. (A) Time SK8 was shorter in groups B and C than in group A; (B) Efficacy SK8 was better in groups B and C than in group A. * $P < 0.05$, ** $P < 0.005$, Tukey multicomparison *post hoc* test.

TABLE 2. CORRELATION AMONG THE REPRESENTATIVE PARAMETERS AND OTHER MOTION ANALYSIS DATA OF LAPMENTOR IN THE 92 PARTICIPANTS

Task	<i>n</i>	Correlated parameters	<i>R</i>	<i>P</i> value
SK3 Time needed to touch 1 ball (Time SK3)		Total No. of motions of right-handed instruments	0.431	0.000
		Total no. of motions of left-handed instruments	0.357	0.000
		Efficacy of movement of left-handed instruments	-0.218	0.037
		Average speed of movement of right-handed instruments	-0.520	0.000
		Average speed of movement of left-handed instruments	-0.546	0.000
SK5 Accuracy in clipping the tube (Accuracy SK5)		Time needed to perform the required task	-0.297	0.004
		Total no. of motions of right-handed instruments	-0.461	0.000
		Total no. of motions of left-handed instruments	-0.313	0.002
		Total length of movement of right-handed instruments	-0.355	0.001
		Total length of movement of left-handed instruments	-0.360	0.000
		Efficacy of movement of right-handed instruments	0.459	0.000
		Efficacy of movement of left-handed instruments	0.303	0.003
SK8 Time needed to coagulate 1 band (Time SK8)		Total no. of motions of right-handed instruments	0.785	0.000
		Total no. of motions of left-handed instruments	0.756	0.000
		Total length of movement of right-handed instruments	0.622	0.000
		Total length of movement of left-handed instruments	0.654	0.000
Efficiency in electrocoagulation (Efficacy SK8)		Total no. of motions of left-handed instruments	0.228	0.029
		Total length of movement of left-handed instruments	0.300	0.004
		Average speed of movement of left-handed instruments	0.309	0.003

physicians were allowed to perform the test only once; hence, there was no warm-up or practice session that may have likewise altered our results.

Usefulness of VR simulators depends on the relationship between the performances on VR simulators and real surgeries (ie, predictive validity). Ahlberg and coworkers⁸ demonstrated better outcome of laparoscopic appendectomy in a pig by students who showed better results in a VR simulator (MIST-VRTM). One of the most impressive results of our study is the demonstration between VS and LapMentor motion analysis data specifically with regard to the accuracy of the performances in SK4, 5, and 8. Physicians who showed accurate performances on a VR simulator demonstrated a safe and accurate laparoscopic procedure as judged by their videotape assessment. The videos assessed in this study were one of the best videos of the participants, just as in the ESSQ system. It is possible that physicians who showed low accuracy on a VR simulator performed inaccurate performances even in their best operations.

FIG. 3. Correlation between video score and Accuracy SK458 of LapMentor basic tasks.

VS in our study also demonstrated negative correlation with the age of physicians and the years of laparoscopic experiences. The interpretation of these results is difficult. One possible interpretation is that the older physicians who started laparoscopic surgeries by themselves in the beginning of the laparoscopic era did not have formal training related to the accurate performance of procedures, resulting in poor results in LapMentor and VS compared with the younger physicians who had more opportunity for a more robust educational experience in laparoscopic surgery. Actually, however, the correlation coefficient was relatively low, and multiple linear regression analysis showed no correlation. Further study is necessary to better test whether there is indeed a strong correlation between performances on VR simulators and real surgical performance.

The difficult issue is how to evaluate a surgeon's performance in actual live surgeries. The ESSQ system, assessing surgical skills on video in a double-blinded fashion, is one of the most objective methods of surgical skills assessment. The results of this study could be reflected by the criteria of video assessment of the ESSQ system. The video assessment guideline of the ESSQ system is mainly dependent on the safe and accurate (ie, error free) performance of the procedure, not on speed of the procedure. It is well understandable that several parameters representing the accuracy of the performance in LapMentor would demonstrate a positive correlation to VS.

A limitation of our study is the accuracy and reproducibility of the video assessment. As stated in our previous report,⁴ the average discrepancy of VS by two referees was 6.0 points in the ESSQ system. Although the video assessment in this study was performed according to the criteria of the ESSQ system by experienced referees of the ESSQ system who had enough training of video assessment, there exists the possibility of some discrepancy among the referees as a nature of this type of skill assessment. Furthermore, only 4 of 40

physicians had $VS < 60$. To determine reliability of the LapMentor, which could discriminate pass (> 60) and fail (< 60) of VS, further study with a larger number of participants would be necessary.

We also studied which parameters of movement of the instruments are related to the parameters representing general quality of performances in LapMentor, such as time or accuracy of tasks. Time SK3 and Time SK8 are correlated with the efficacy of movement of instruments. In particular, Time SK8 showed strong correlation to total number of motions of instruments, whereas speed of movement of instruments had no effects. Speed of movement of instruments showed moderate correlation to Time SK3, which is a simpler task compared with SK5 or SK8. Accuracy SK5 also had moderate correlation to quality of movement of instruments, whereas speed of movement of instruments showed no correlation. Correlation between Efficacy SK8 and motion analysis data are difficult to interpret because correlation coefficients are too small. The correlation between Efficacy SK8 and the movement of the left-handed forceps may indicate the importance of movement of the nondominant hand. Although we have no data with regard to the dominant hand of the participants, the percentage of left-handed men was reported to be 8% to 15%.⁹ Our results indicate that smooth and ideal movement of instruments is more important than speed of the movement of instruments to achieve good performances.

Conclusion

Our study demonstrates the usefulness of VR simulators in the evaluation of laparoscopic surgical skills and the ability of the VR simulator to correlate with actual surgical laparoscopic performance. This is a very preliminary step, however, and much work needs to be done to reliably evaluate the proficiency level needed for surgeons to safely perform invasive laparoscopic procedures in the clinical realm.

Acknowledgment

This work was supported in part by Grant-in-Aid 19591875 from the Ministry of Education, Science, Sports and Culture, Japan.

Disclosure Statement

No competing financial interests exist.

References

1. Clayman RV, Kavoussi LR, Soper NJ, et al. Laparoscopic nephrectomy. *N Engl J Med* 1991;324:1370–1371.
2. Go H, Takeda M, Takahashi H, et al. Laparoscopic adrenalectomy for primary aldosteronism: A new operative method. *J Laparoendosc Surg* 1993;3:455–459.
3. Scott DJ, Valentine RJ, Bergen PC, et al. Evaluating surgical competency with the American Board of Surgery In-Training Examination, skill testing, and intraoperative assessment. *Surgery* 2000;128:613–622.
4. Matsuda T, Ono Y, Terachi T, et al. The endoscopic surgical skill qualification system in urological laparoscopy: A novel system in Japan. *J Urol* 2006;176:2168–2172.
5. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005;241:364–372.
6. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. *Br J Surg* 2004;91:1549–1558.
7. McDougall EM, Corica FA, Boker JR, et al. Construct validity testing of a laparoscopic surgical simulator. *J Am Coll Surg* 2006;202:779–787.
8. Ahlberg G, Heikkinen T, Iselius L, et al. Does training in a virtual reality simulator improve surgical performance? *Surg Endosc* 2002;16:126–129.
9. Hardyck C, Petrinoich LF. Left-handedness. *Psychol Bull* 1977;84:385–404.

Address correspondence to:

Tadashi Matsuda, M.D.
Department of Urology and Andrology
Kansai Medical University
Shin-machi 2-3-1, Hirakata
Osaka 573-1191
Japan

E-mail: matsudat@takii.kmu.ac.jp

Abbreviations Used

ESSQ = Endoscopic Surgical Skill Qualification
SK = skill task
VR = virtual reality
VS = videotape score