Objective Surgical Skill Assessment: An Initial Experience by Means of a Sensory Glove Paving the Way to Open Surgery Simulation?

Giovanni Saggio, Eng, PhD,* Alessandra Lazzaro, MD, PhD,† Laura Sbernini, Eng,†
Francesco Maria Carrano, MD,† Davide Passi, Eng,* Arianna Corona, Eng, PhD,†
Valentina Panetta, MSc,‡ Achille L. Gaspari, MD,† and Nicola Di Lorenzo, MD, PhD†

*Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy; †Department of Experimental Medicine and Surgery, University of Rome Tor Vergata, Rome, Italy; and ‡Biostatistics office, L'altrastatistica srl-Consultancy & Training, Rome, Italy

INTRODUCTION: Simulation and training in surgery are very promising tools for enhancing a surgeon's skill base. Accurate tracking of hand movements can be a strategy for objectively gauging a surgeon's dexterity, although "open" work is much more difficult to evaluate than are laparoscopic tasks. To the authors' knowledge, a system taking into account the movements of each finger joint has never been applied to open surgery simulation. This work intends to make up for this shortcoming and to perform a data analysis of the surgeon's entire gesture.

MATERIALS AND METHODS: The authors developed a sensory glove to measure flexion/extension of each finger joint and wrist movement. Totally 9 experts and 9 novices performed a basic suturing task and their manual performances were recorded within 2 days of measurements. Intraclass correlation coefficients were calculated to assess the ability of the executors to repeat and reproduce the proposed exercise. Wilcoxon signed-rank tests and Mann-Whitney *U*-tests were used to determine whether the 2 groups differ significantly in terms of execution time, repeatability, and reproducibility. Finally, a questionnaire was used to gather operators' subjective opinions.

RESULTS: The experts needed a similar reduced execution time comparing the 2 recording sessions ($p = 0.09$), whereas novices spent more time during the first day ($p = 0.01$). Repeatability did not differ between the 2 days, either for experts ($p = 0.26$) or for novices ($p = 0.86$). The 2 groups performed differently in terms of time ($p < 0.001$), repeatability ($p = 0.01$), and reproducibility ($p < 0.001$) of the same gesture. The system showed an overall moderate repeatability (intraclass correlation coefficient: experts = 0.64; novices = 0.53) and an overall high reproducibility. The questionnaire revealed performers' positive feedback with the glove.

CONCLUSIONS: This initial experience confirmed the validity and reliability of the proposed system in objectively assessing surgeons' technical skill, thus paving the way to a more complex project involving open surgery simulation. (J Surg Ed 72:910-917. © 2015 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: skill, assessment, education, open surgery, sensory glove

COMPETENCIES: Medical Knowledge, Practice-Based Learning and Improvement, Systems-Based Practice, Professionalism

INTRODUCTION

Despite the enormous improvements in effectiveness of surgical treatments in the recent years, the criteria for evaluating the surgical skills of trainees remain mainly subjective. In fact, using observation and experience, expert examiners continue to be the judges of the learners' skills. As the education and training of surgeon still remains a matter of "learning on the job," the apprenticeship learning model is mainly based on observation, imitation, and instructions.

Meanwhile, with the dramatic changes introduced into clinical practice, advances in modern medical and surgical practice have come to be associated with meaningful changes in medical education. Consequently, assessment of learning levels should be based on structured methods and objectivity.[1]

Correspondence: Inquiries to Giovanni Saggio, Eng, PhD, Department of Electronic Engineering, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy; fax: (062) 331-4067; E-mail: saggio@uniroma2.it, wmdtg@email.it

1931-7204/$30.00
http://dx.doi.org/10.1016/j.jsurg.2015.04.023

Because of these innovations, new education methods are being introduced, including basic surgical maneuvers to be practiced by trainees on models and simulators before working in an actual operating room. As Satava[2] predicted, the surgical residents should repeatedly practice surgical procedures until they are perfect before performing surgery on patients. Therefore the "learning curve" (with all the possible mistakes involved) takes place in the laboratory, rather than being inflicted on the patient, leading, of course, to a dramatic improvement in patient safety.[3]

In such a scenario, objective assessment is mandatory, especially as deficiencies in training and performance are difficult to correct without an objective score system. Since 1990s, different systems have been developed to simulate operation procedures and to allow objective assessment without having to bring in expert observers; in this regard, dexterity analysis systems, motion analysis, and virtual reality have all been introduced.[4]

Simulation in surgery, from low-fidelity bench-top models to high-fidelity technologically advanced virtual reality systems,[5,6] has achieved a widespread acceptance, especially in minimally invasive surgery learning.[7] Of late, efforts have been mainly devoted to the laparoscopic operations, for which bidimensional electronic-mediated images and fulcrum-linked motion of the instruments can be easily reproduced for simulation purposes. Despite continual advances in minimally invasive surgery simulators, simulation and objective assessment in open surgery remains a critical component in modern surgical education.[8,9] At present, simulation of open techniques is more challenging and spatially complex, as a comprehensive and immersive environment is needed.[10]

Consequently, assessment in open surgery is still performed via experts' judgement based on observation and verbal feedback or structured assessment,[8] such as the most validated "Objective Structured Assessment of Technical Skill" (OSATS).[11,12] As an alternative, tracking the surgeons' hands, e.g., can greatly improve open surgery simulation and training. Acquisition of surgical hand maneuvers by means of an automated measuring system can provide objective data for assessing real surgical skill in the operating room itself.[9,13]

With the aim of filling the described lack of open surgery skill assessment, we propose an innovative system based on a sensory glove capable of acquiring, storing, and analyzing up to 17 degrees of freedom of the hand in practicing surgical gestures. This is part of a more complex ongoing project where arms and trunk posture and motion can all be measured by means of sensorized garments, leading to quantitative knowledge of the surgeon's "actions." The use of a classification strategy and a virtual reality environment would provide additional data. Overall data gathered from experts' performances can be further implemented, resulting in an average gesture that can be the reference standard for novices and residents.

We present the features related to the initial experience and the validation of an objective assessment tool suitable for open surgery simulation and training that could be easily implemented into standard practice. We focus on evaluating manual performance during surgeons' training in open surgery through measuring his/her hands executing a basic surgical task in a simulated bench operation with real surgical instruments. We compare a group of expert surgeons with respect to a group of novices, our assumption being that skill could lie in the organization of motion[13] and that certain variables, such as execution time and gesture repeatability and reproducibility, can provide significant information for differentiating between surgeons' technical skill levels.

MATERIALS AND METHODS

In order to measure the hand gestures, we developed a sensory glove (Fig. 1) based on acquired experiences for nonsurgical purposes.[14-18] The glove was equipped with 14 flex sensors (by Flexpoint Sensor Systems, Inc., Draper, UT) and a 3-axis accelerometer (ADXL335, by Analog Devices, Inc., Norwood, MA) able to measure the flex/extension capabilities of the finger joints of a human hand, plus the wrist movements. Flex sensors were placed on distal interphalangeal, proximal interphalangeal, and metacarpophalangeal finger joints, and the accelerometer was placed on backside of the hand (Fig. 2). In total 17 signals were collected by means of a custom-made prototype board connected to a computer. The accuracy and repeatability of the measures obtained with this glove are similar to those of others reported in the literature.[19,20]

We realized gloves in 2 sizes (small and medium) so as to best fit the hands of the testers, these consisting of 2 groups of volunteers. There were 9 "experts" (3 men and 6 women, aged 33-35), i.e., surgical residents in their final 3 years of training and with a high level of expertise regarding the task

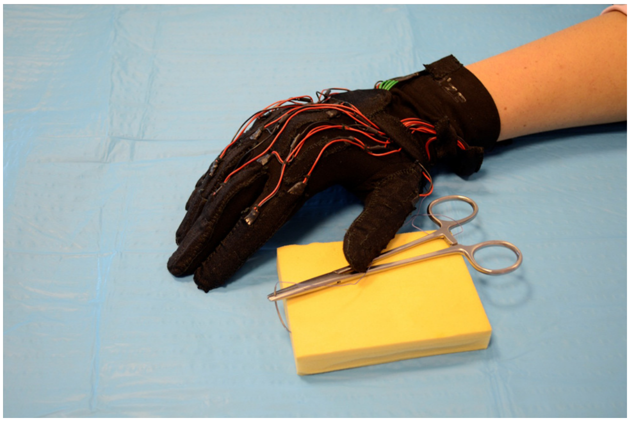
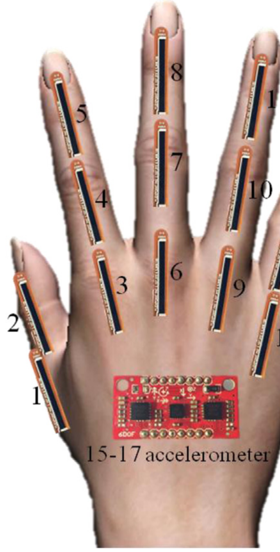


FIGURE 1. The sensory glove was provided with 14 flex sensors and a 3-axis accelerometer able to measure the flex/extension capabilities of the finger joints and the wrist movements.



#	Sensor
1	1-MCPJ
2	1-IPJ
3	2-MCPJ
4	2-PIPJ
5	2-DIPJ
6	3-MCPJ
7	3-PIPJ
8	3-DIPJ
9	4-MCPJ
10	4-PIPJ
11	4-DIPJ
12	5-MCPJ
13	5-PIPJ
14	5-DIPJ
15	accelerometer x-axis
16	accelerometer y-axis
17	accelerometer z-axis

FIGURE 2. Sensors placement and nomenclature: 14 flex sensors were placed on distal interphalangeal joints (DIPJ), proximal interphalangeal joints (PIPJ), and metacarpophalangeal joints (MCPJ), and a 3-axis (x, y, and z) accelerometer was placed on the hand's backside. Fingers nomenclature: 1 (thumb), 2 (index), 3 (middle), 4 (digitus annularis), and 5 (fifth digit).

performed in this study, and 9 "novices" (7 men and 2 women, aged 26-30), i.e., inexperienced engineering students and research engineers without previous training in surgical tasks. An informed consent was signed by individual participants before the study.

The performed task consisted in a simple stitch (a single interrupted suture) with a double-needle passage across a vertical incision, previously made on a soft tissue pad, to simulate the human skin. No attempt to evaluate knot-tying gestures was performed at the end of the suture.

We adopted a standard protocol to perform the test, following the same conditions and procedure throughout. More specifically, we asked each subject to sit down with the arm forming a 90° angle to a front desk, the dominant hand wearing the glove with a neutral position defined by reference lines on the desk, and the soft tissue always with a fixed place. Each task started with the performer's hand holding the needle holder and concluded in the same neutral position. Each participant attested to being right handed and performed the tests using his/her dominant hand.

Each operator repeated the task 10 times, and the overall procedure was repeated in 2 consecutive days. Thus, we collected data of 20 dynamic measurements for each subject, from 2 groups (precisely, 10 for the first day and 10 for the second).

We decided not to map the analog sensor values to the corresponding joint angles, thus bypassing the phase of glove calibration. This phase is time consuming because an accurate calibration of the glove was needed for each new user; it could also introduce several errors due to the conversion of the analog values into angles.[19,21] In this manner, each task recording was made of the 17 raw sensor readings, namely the analog temporal waveforms that indirectly described hand joint angles.

A data-cutting algorithm, based on a discrete moving average filter, was applied so as to discard possibly insignificant information in the recorded data, as some temporal samples may be registered before and after the meaningful movements. To determine the execution time of each repetition, we flagged the start and the end sample of each significant gesture. For each subject we considered the median time value among the 10 repetitions of each measurement session and then a unique median value between the 2 days as a reference time.

Raw data were filtered and smoothed with a moving average filter (window size of 5 samples) to remove any spurious spikes in the data, and then auto-scaled. To compare waveforms with different execution times, data were time-normalized by means of a resampling procedure. In this manner, N uniformly spaced time points represented each sensor's temporal waveform, losing information regarding time but maintaining motion characteristics.

At the end of the preprocessed phase, data of each repetition were arranged in a matrix $N \times n$, where the n columns were the 17 sensor's waveforms and where N was set at 1000.

The intraclass correlation coefficient (ICC)[22] was used to obtain system reliability and to assess the executors' ability to both repeat and reproduce the exercise proposed. A repeatability analysis was performed within the 10 consecutive repetitions in each measurement session and the coefficient computed by comparing the waveforms separately for each sensor. The median value among the 17 obtained was calculated to generalize the result to the whole gesture. The ICC was estimated for each subject separately for the 2 sessions; we then took into account a unique average value of the 2 days for each subject as a reference value. For the reproducibility analysis, we considered the mean value of the 10 measures collected in a single day whereas the ICC was computed with these average results for each sensor. Again, we considered for each subject the median value of the coefficients among all sensors to match the results to the whole gesture.

The aim of the statistical analysis was to identify significant differences by comparing both results within the same group and results between the 2 groups; p-values less than 0.05 were considered statistically significant.

Separate analysis was conducted for the experts and the novices ("within" group analysis) using the Wilcoxon signed-rank test for paired samples; its aim was to evaluate statistical differences between the 2 measurement sessions in terms of execution time and repeatability.

Mann-Whitney *U*-tests for independent samples between overall unique results confirmed the possibility of differentiating the subjects' ability through the collected values of time, ICCs for repeatability, and ICCs for reproducibility ("between" groups analysis).

TABLE. User Feedback Questionnaire*

Questionnaire Items	Mean Score†	Total Score‡
1 I felt comfortable as the glove was put on	5.4 (1.0)	5.3 (0.5)
2 I did not feel my fingers were put into any uncomfortable position as the glove was put on	5.4 (1.0)	
3 I did not feel any restriction to movement with this glove	5.1 (1.2)	
4 I felt comfortable performing the activities in this study with this glove (I had no trouble using surgical instruments)	5.8 (0.9)	
5 The glove did not feel too tight (it did not make my hands or fingers tingle)	5.6 (0.9)	
6 I feel like I can bend my fingers just like I can without wearing the glove	5.0 (1.1)	
7 The glove did not feel too hot or too cold	6.2 (1.2)	
8 I did not feel a reduction in tactile sensitivity of the fingers with this glove	4.4 (1.5)	
9 I feel I can do most surgical tasks while wearing this glove	5.3 (1.0)	
10 I felt a reduction in tactile sensitivity of the fingers and this has not compromised the quality of surgical task	4.5 (1.7)	
11 I felt comfortable as the glove was removed	5.4 (1.4)	
12 I did not feel like my fingers were put into any uncomfortable position as the glove was removed	5.7 (1.5)	

SD, standard deviation. Adapted from Gentner and Classen[20] and Simone et al.[23]

*Answers range from 1 to 7 (strongly disagree, disagree, somewhat disagree, neutral, somewhat agree, agree, and strongly agree).

†Data are presented as mean (SD) score per item among all subjects (9 experts and 9 novices).

‡Data are presented as mean (SD) of the total score among the 12 items.

Data were obtained and statistical analysis was performed, respectively, using MATLAB (MATLAB 2013a, The MathWorks, Inc., Natick, MA) and SPSS software (IBM SPSS Statistics, v. 20.0, IBM Corp., Armonk, NY), and power analysis was performed using the PASS software (PASS 13 NCSS, LLC. Kaysville, UT).

At the end of the protocol, we administered a short questionnaire (Table), to elicit feedback on the comfort of donning/doffing the glove and of performing the simulated surgical task. For each item of the questionnaire that we asked participants to answer, scaling was from 1 to 7, ranging from complete disagreement to complete agreement (strongly disagree, disagree, somewhat disagree, neutral, somewhat agree, agree, and strongly agree).

RESULTS

Figure 3A concerns the time required to perform the test for each tester. Data demonstrated no significant differences within the group of experts between the 2 days (Day 1: median = 6.20 s [interquartile range {IQR}: 4.78-6.57]; Day 2: median = 5.56 s [IQR: 4.85-5.75]; $p = 0.09$; Fig. 3A); conversely, we registered significant differences within the group of novices (Day 1: median = 11.52 s [IQR: 10.10-12.67]; Day 2: median = 8.00 s [IQR: 7.03-10.19]; $p = 0.01$; Fig. 3A). Novices spent more time in completing the task in the first day and less time during the second day. This can be easily explained considering that the first day was a sort of "training session" to prepare for the second day of tests.

The time necessary to complete the exercise was significantly different between groups ($p < 0.001$), with experts spending a shorter time (median = 5.77 s [IQR: 4.77-6.34]) with respect to novices (median = 10.35 s [IQR: 8.21-10.89]).

Repeatability analysis provided results for each participant, as shown in Figure 3B. From such data we could not differentiate the 2 measurement sessions by means of the ICC values for both groups (experts group $p = 0.26$, novices group $p = 0.86$). The gesture can be similarly repeatable even if the performer is less experienced and the gesture is not necessarily executed correctly. In particular, as shown in Figure 3B, for the experts group the median repeatability values were 0.69 (IQR: 0.53-0.72) for the first day and 0.68 (IQR: 0.64-0.75) for the second. For the novices the median repeatability values were 0.55 (IQR: 0.48-0.60) for the first day and 0.58 (IQR: 0.46-0.63) for the second one. The overall data demonstrated that experience was statistically significant in terms of repeatability ($p = 0.01$), median ICC = 0.64 (IQR: 0.58-0.76) for the experts and ICC = 0.53 (IQR: 0.47-0.57) for the novices.

Reproducibility analysis (Fig. 3B) proved that the 2 groups could be differentiated in terms of reproducibility of gesture ($p < 0.001$) and clearly showed how a surgical gesture, although repeatable, may not be performed in a reproducible manner without the required experience. In particular the median reproducibility value of experts was 0.91 (IQR: 0.86-0.92), whereas for novices it was 0.69 (IQR: 0.69-0.77; Fig. 3B).

Finally, the user feedback questionnaire (Table) revealed that performers had a positive feedback with the sensory glove, the total mean score = 5.3 (standard deviation = 0.5) for all participants.

DISCUSSION

Improving surgical dexterity is mandatory for surgeons in training to perform all tasks better, from simple wound

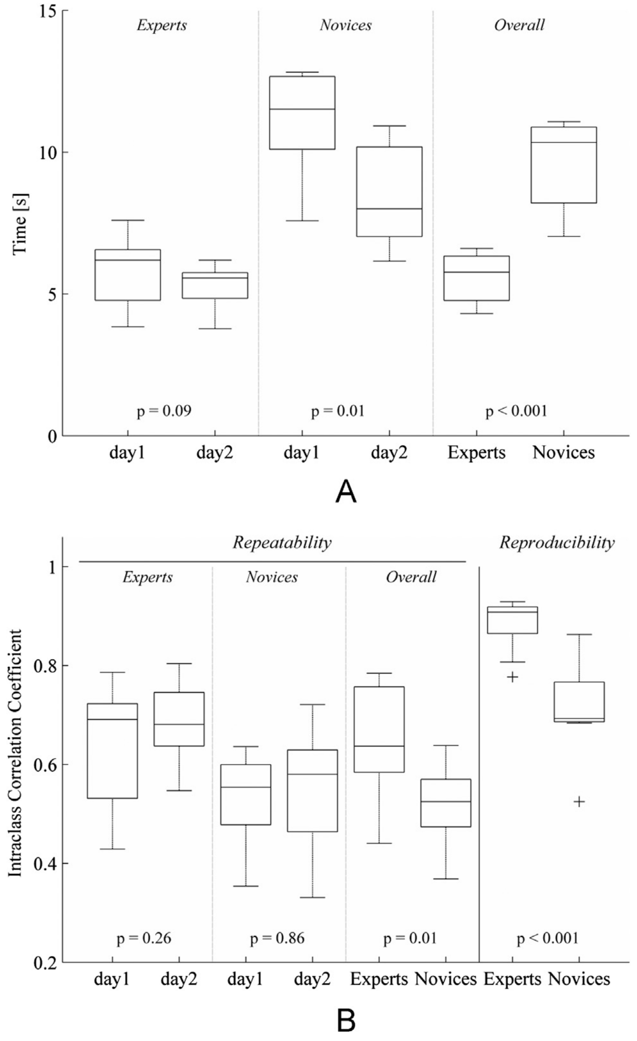


FIGURE 3. Box plot for the statistical tests data and relative p values. Participants are divided into 2 groups: first of 9 "experts" and a second of 9 "novices." (A) Time analysis; (B) repeatability (left) and reproducibility (right) analysis. Statistical significance is defined as $p < 0.05$.

closure to highly complex therapeutic procedures. New approaches, in combination with new technologies such as surgical simulators, objectively evaluate surgical skills and can fundamentally improve surgical education and training. Competency assessment based upon quantifiable criteria measures should replace the traditional subjective one.[24] Moreover, as a general principle, repetitive performance of a specific task results in a much better performance than does general-purpose training.[25] Accordingly, measurement of hand movements can be fundamental to objectively evaluate surgical dexterity, and offers a valid approach for remedying the current lack of experience in open surgery simulation and training.

Several efforts have been made to use the trainees' gestures measure to quantitatively evaluate their skill. The involved technologies used electromagnetic sensors placed on dorsum of each hand during the performance, e.g., the Imperial College Surgical Assessment Device,[26,27] as well as optical tracking systems, magnetic tracking technologies, and video graphic recording systems.[28,29] Methodologies based on tracking surgical movements with inertial measurement units have also been proposed to quantify surgical performance.[30]

The aforementioned methodologies are all suitable both for minimally invasive and open surgery training. However, they present the following limitations:

- they are potentially sensitive to surrounding noise, consequently needing a controlled environment;
- they are potentially compromised by the line of sight interruption between sensors and cameras (objects and self-occlusion);
- the hand is considered as a single point in the space, whereas in open surgery it is important to evaluate the fine movements of the surgeon's fingers freely manipulating the instruments.

Shifting from passive external tracking of movements to active registration of hand motion can simplify simulation and training. A substantial difference between open and minimally invasive surgery procedures is the involved hand's degrees of freedom. In open surgery we must track almost all the degrees of freedom of each hand, not just the restricted 4 degrees of freedom of minimally invasive surgery.[13] Accordingly, gloves with embedded sensors, termed sensory gloves, can solve the mentioned drawbacks. Such sensory gloves, when worn by the surgeon, track his/her movements, giving both global information for the hands and local information for the joints of each finger.

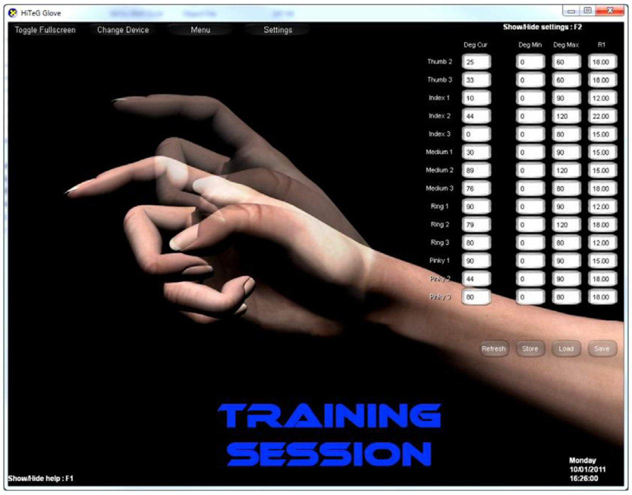
We are aware of very few assessing methods based on sensory gloves reported in literature and they are applied to minimally invasive surgery or microsurgery training.[31-33] To our knowledge, this is the first study applied to open surgery simulation that objectively assesses surgical skills through a system able to record movements of the joints of each single finger. It describes the initial experience with the use of a sensory glove as a training tool in open surgery, comparing a group of expert surgeons to a group of inexperienced subjects.

Results show significant differences between the 2 groups and confirm that the developed system respects some of the basic principles of assessment about reliability, validity, and feasibility.[1,34-36] Based on a landmark study[37] and following common practice, we can empirically consider the repeatability and reproducibility as "poor to fair" when $\text{ICC} < 0.40$, "moderate" when $0.41 < \text{ICC} < 0.60$, "good" when $0.61 < \text{ICC} < 0.80$, and "excellent" when $\text{ICC} > 0.81$. Accordingly, the experts show a good repeatability of their surgical gestures, whereas novices globally show only a moderate repeatability. The resulting overall moderate repeatability considering all participants, together with the high values of reproducibility ("excellent" for the experts and "good" for the novices), demonstrates the reliability of the evaluation instrument (test-retest reliability) and validates its potential for objectively assessing testers' performance. The outcomes, coming from the "between" statistical analysis, demonstrate that the 2 different surgical expertise levels are distinguished, by assessing the proposed system's construct validity.

Moving on to additional considerations, the positive result of the questionnaire highlights the participants' acceptance of the glove. It is also worth noticing that the system is portable and quite user-friendly, to allow learners to practice anywhere, even outside the clinical setting, these features being fundamental for regular practice. Furthermore, the developed prototype (glove and electronics) is cost-effective, as the price of the components is in the hundreds of dollars range. The glove, in addition to the computer, might be an inexpensive alternative to several commercially available devices, priced well beyond 2000 dollars.[19] and currently not for sale, but we can assume that the commercial cost would be lower with the increase in demand.

Further developments would include modification of the glove for sterilization purposes and for surgical use in real procedures, as well as subgroup analysis (masters, experts, intermediates, residents of last or first year, etc.) to discriminate even more specifically the skill progression. The one described in this article is an initial set of experiments: limitations can arise as we assessed only a single low-complexity task and enrolled a small number of subjects. A small sample size could affect the results of the statistical tests, as the power of the study is not adequate for all the performed tests. Considering the generally accepted power threshold of 80%, the problem has adequate power for 2 "between" tests (time and reproducibility analysis) and for a "within" test (novices regarding the metric of time).

Our future work would add several additional tasks and include more participants, in order to have a more powerful statistics and a more comprehensive skill evaluation during surgical procedures.



	Deg Cur	Deg Min	Deg Max	R1
Thumb 2	25	0	60	18.00
Thumb 3	33	0	60	18.00
Index 1	10	0	90	12.00
Index 2	44	0	120	22.00
Index 3	0	0	80	15.00
Medium 1	30	0	90	15.00
Medium 2	89	0	120	15.00
Medium 3	76	0	80	18.00
Ring 1	90	0	90	12.00
Ring 2	79	0	120	18.00
Ring 3	80	0	80	12.00
Pinky 1	90	0	90	15.00
Pinky 2	44	0	90	18.00
Pinky 3	80	0	80	18.00

Monday
10/01/2011
16:26:00

FIGURE 4. The virtual reality environment with the avatar hand representation. A training session with a ghost avatar of the learner's hand superimposed on the reference avatar hand of an expert surgeon.

The aim being to improve this expertise recognition, it seems evident that this could be evaluated at a higher statistical level. To this end, a method based on advanced classification algorithms is more appropriate than a low-level statistical analysis. Consequently, we are currently introducing automatic classification methods, such as artificial neural networks, to further improve our analysis.

To enhance our virtual reality environment, we developed a customized avatar that allows both real-time and offline representations of the recorded movements as well as the further possibility of superimposing a ghost avatar of the learner on a reference avatar of an expert surgeon (Fig. 4). Novices can thus repeat a simulated surgical task as many times as they want, comparing their performance to previous training sessions and to a large database of experts' gestures. Learners are thus able to self-evaluate their skill rehearsals, continuing, if necessary, the learning cycle until they achieve an acceptable skill level, without the physical presence of an expert as a tutor.

CONCLUSIONS

As discussed, our initial experience confirmed the validity and the reliability of the proposed method of evaluating open surgical gesture, although further investigations are still required. The obtained data would shed new light on objective evaluation of attending the surgeon's learning curve, scoring it, and comparing trainee's expertise levels to those of master surgeons, paving the way to faster acquisition by means of electronic help.

ACKNOWLEDGMENTS

The authors thank all subjects who kindly volunteered and participated in this study.

REFERENCES

1. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg*. 2010;97(7):972-987.
2. Satava RM. Virtual reality surgical simulator. The first steps. *Surg Endosc*. 1993;7(3):203-205.
3. Satava RM. The revolution in medical education—the role of simulation. *J Grad Med Educ*. 2009;1(2):172-175.
4. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *Br Med J*. 2003;327(7422):1032-1037.

5. de Montbrun SL, Macrae H. Simulation in surgical education. *Clin Colon Rectal Surg*. 2012;25(3):156-165.

6. Saggio G, Pinto CA. Virtuality supports reality for e-health applications. In: Jae-Jin Kim, ed. *Virtual Reality*; *InTech Publications*; 2011:247-272.

7. Dankelman J, Di Lorenzo N. Surgical training and simulation. *Minim Invasive Ther Allied Technol*. 2005;14(4):211-213.

8. Fonseca AL, Evans LV, Gusberg RJ. Open surgical simulation in residency training: a review of its status and a case for its incorporation. *J Surg Educ*. 2013;70(1):129-137.

9. Davies J, Khatib M, Bello F. Open surgical simulation–a review. *J Surg Educ*. 2013;70(5):618-627.

10. Di Lorenzo N. Ideal V.R. systems: is there a 'holy grail' in simulation system land? In: Satava RM, Gaspari A, Di Lorenzo N, eds. *Emerging Technologies in Surgery*. Berlin Heidelberg: Springer, 2007: 47-55.

11. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.

12. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *Am J Surg*. 1997;173(3):226-230.

13. Reiley CE, Lin HC, Yuh DD, Hager GD. Review of methods for objective surgical skill evaluation. *Surg Endosc*. 2011;25(2):356-366.

14. Saggio G. Mechanical model of flex sensors used to sense finger movements. *Sens Actuators A Phys*. 2012;185:53-58.

15. Saggio G, Bocchetti S, Pinto CA, Orengo G. Wireless Data Glove System Developed for HMI. International Symposium on Applied Sciences in Biomedical and Communication Technologies. Isabel; 2010;1-5.

16. Saggio G, De Sanctis M, Cianca E, Latessa G, De Santis F, Giannini F. Long Term Measurement of Human Joint Movements for Health Care and Rehabilitation Purposes. Proceedings for the 2009. 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology; Aalborg, Denmark. 2009;674-678.

17. Saggio G, Bizzarri M. Feasibility of Teleoperations with Multi-Fingered Robotic Hand for Safe Extravehicular Manipulations. *Aerospace Sci Technol*. 2014;39:666-674.

18. Saggio G, Bianchi L, Castelli S, Santucci MB, Fraziano M, Desideri A. In vitro analysis of pyrogenicity and cytotoxicity profiles of flex sensors to be used to sense human joint postures. *Sensors*. 2014;14(7):11672-11681.

19. Dipietro L, Sabatini AM, Dario P. A survey of glove-based systems and their applications. *IEEE Trans Syst Man Cybern C Appl Rev*. 2008;38(4):461-482.

20. Gentner R, Classen J. Development and evaluation of a low-cost sensor glove for assessment of human finger movements in neurophysiological settings. *J Neurosci Methods*. 2009;178(1):138-147.

21. Heumer G, Ben Amor H, Weber M, Jung B. Grasp recognition with uncalibrated data gloves—a comparison of classification methods. *Proc IEEE Virtual Real Conf*. 2007:19-26.

22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.

23. Simone LK, Sundarrajan N, Luo X, Jia Y, Kamper DG. A low cost instrumented glove for extended monitoring and functional hand assessment. *J Neurosci Methods*. 2007;160(2):335-348.

24. Satava RM. Disruptive visions: surgical education. *Surg Endosc*. 2004;18(5):779-781.

25. Di Lorenzo N, Coscarella G, Faraci L, Konopacki D, Pietrantuono M, Gaspari AL. Robotic systems and surgical education. *JSLS*. 2005;9(1):3-12.

26. Darzi A, Mackay S. Assessment of surgical competence. *Qual Health Care*. 2001;10(suppl 2):ii64-ii69.

27. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg*. 2001;193(5):479-485.

28. Sonnadara R, Rittenhouse N, Khan A, et al. A novel multimodal platform for assessing surgical technical skills. *Am J Surg*. 2012;203(1):32-36.

29. Rittenhouse N, Sharma B, Sonnadara R, Mihailidis A, Grantcharov T. Design and validation of an assessment tool for open surgical procedures. *Surg Endosc*. 2014;28(3):918-924.

30. Watson RA. Quantification of surgical technique using an inertial measurement unit. *Simul Healthc*. 2013;8(3):162-165.

31. King RC, Atallah L, Lo BP, Yang GZ. Development of a wireless sensor glove for surgical skills assessment. *IEEE Trans Inf Technol Biomed*. 2009;13(5):673-679.

32. Kahol K, Krishnan NC, Balasubramanian VN, Panchanathan J, Smith M, Ferrara J. Measuring

movement expertise in surgical tasks. MULTIMEDIA '06 Proceedings of the 14th Annual ACM International Conference on Multimedia. New York, NY. 2006; 719-722.

33. Amirouche F, Martin JR, Gonzalez M, Fergusson L. Experimental set-up and sensory glove interface for microsurgery. *Proc Inst Mech Eng H*. 2008;222(1):89-99.

34. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc*. 2003;17(10):1525-1529.

35. Carter FJ, Schijven MP, Aggarwal R, et al. Work group for evaluation and implementation of simulators and skills training programmes. Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc*. 2005;19(12):1523-1532.

36. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Assessment of technical surgical skills. *Eur J Surg*. 2002;168(3):139-144.

37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1): 159-174.