

Development of force-based metrics for skills assessment in minimally invasive surgery

Ana Luisa Trejos · Rajni V. Patel ·
Richard A. Malthaner · Christopher M. Schlachta

Received: 24 June 2013 / Accepted: 12 January 2014 / Published online: 12 February 2014
© Springer Science+Business Media New York 2014

Abstract

Background The loss of haptic information that results from the reduced-access conditions present in minimally invasive surgery (MIS) may compromise the safety of the procedures. This limitation must be overcome through training. However, current methods for determining the skill level of trainees do not measure critical elements of skill attainment. This study aimed to evaluate the usefulness of force information for the assessment of skill during MIS.

Methods To achieve the study goal, experiments were performed using a set of sensorized instruments capable of measuring instrument position and tissue interaction forces. Several force-based metrics were developed as well as metrics that combine force and position information.

Results The results show that experience level has a strong correlation with the new force-based metrics presented in this article. In particular, the integral and the

derivative of the forces or the metrics that combine force and position provide the strongest correlations.

Conclusions This study showed that force-based metrics are better indications of performance than metrics based on task completion time or position information alone. The proposed metrics can be automatically computed, are completely objective, and measure important aspects of performance.

Keywords Force sensing · Skills assessment · Performance metrics · Minimally invasive surgery · Sensorized instruments

The acquisition of technical skill in performing a difficult task is a complex phenomenon in which the time it takes to become proficient is a function of many variables, including the amount of practice, the kind of practice, and the type of feedback provided [1]. This is especially true for the surgical skills required to perform minimally invasive surgery (MIS), a type of procedure in which the surgery is performed through small incisions using long, slender instruments that pivot about the incision point.

With the introduction of MIS, the rise in the number of surgery-related injuries created an awareness of the need to measure technical competence properly [2] and to develop better training methods. Achieving technical competence in MIS procedures is not easy. The learning process is affected by many perceptual and motor limitations that steepen the learning curve [3–5]. This has led to the introduction of new requirements for training, in which all residents need to achieve certain competency levels before operating on humans [5].

Thanks to new technologies that allow for the recording of activities during training, it currently is possible to

Rajni V. Patel is the Project leader.

A. L. Trejos · R. V. Patel · R. A. Malthaner · C. M. Schlachta
Canadian Surgical Technologies and Advanced Robotics,
Lawson Health Research Institute, 339 Windermere Road,
London, ON, Canada

A. L. Trejos (✉) · R. V. Patel
Department of Electrical and Computer Engineering, The
University of Western Ontario, London, ON, Canada
e-mail: atrejos@uwo.ca

R. V. Patel · R. A. Malthaner · C. M. Schlachta
Department of Surgery, The University of Western Ontario,
London, ON, Canada

R. A. Malthaner
Department of Epidemiology and Biostatistics, The University
of Western Ontario, London, ON, Canada

develop metrics and quantitative descriptors that characterize technical performance. However, identifying how best to quantify technical performance is the subject of extensive research, and no ideal solution has been found.

To understand how motor skills are acquired, Fitts and Posner proposed three stages of motor skill development [6]: the cognitive phase, which is what students do in class (read, watch, and listen); the integrative phase, in which students start to apply the knowledge with some guidance but with a lack of fluidity; and finally, the automatous phase, in which fully independent learning occurs with no supervision or guidance.

Recent studies investigating goal-directed aiming provide deeper insight into how motor skills are developed and how they are affected by extended practice schedules. Woodworth's two-component model shows that goal-directed aiming is composed of two phases: the preprogrammed phase (initial adjustment to bring the limb into the vicinity of the target) and the homing phase (small adjustments required to reach the target accurately) [7].

Many variations on this model have been proposed to reflect more accurately the intricacies of skilled limb control and the effects of practice [7, 8]. In particular, it has been shown that practice has two effects: (1) trainees learn to adapt the initial impulse by accelerating sooner and more aggressively in order to position the limb closer to the target faster, and (2) trainees are able to correct aiming errors faster during the homing phase [7]. Limb motion is controlled directly by adjusting muscular forces. Variability in the muscular forces increases proportionally with the muscular forces required to move the limb [8].

Based on the aforementioned factors, a multiple-process model of limb control is presented by Elliott et al. [8], which builds on Woodworth's two-component model and its variants. In this model, the first component refers to a planning phase that requires the person to optimize the speed and acceleration of the limb (by adjusting muscular forces) in order to place the limb near the target and a corrective component activated late in the movement to correct for differences between the limb and the target position. As trainees learn to control muscular forces better through practice, spatial variability decreases, resulting in greater accuracy at faster speeds. However, the learning process requires trainees to try various things and sense what it feels like to achieve different outcomes at each attempt [9].

As learning occurs by doing, the most important variable for skill acquisition is for how long the trainee practices [10]. However, trainees learn at different rates, highlighting the need to identify objectively the level of a trainee's acquired technical competence. Unfortunately, motor skills are not easy to measure [6, 11], and considerable controversy exists regarding the best method for assessing motor skills.

Validated assessment methods

The standard method of assessing motor skills in health education is through the use of checklists or standard rating scales. Global rating scales (GRS) in general have been proposed for use in many areas [12, 13]. A more standardized method of laboratory training is called the objective structured assessment of technical skill (OSATS), which combines checklists and GRS to provide a structured evaluation that attempts to be objective and readily accessible and allows the measurement of a proper learning curve [6, 12]. These performance metrics have the following limitations: they are subjective [11, 14, 15]; they provide no feedback during the learning of complex skills [16]; they are not trainee or procedure specific [13]; and they require extra cost and time due to the need for an evaluator [11].

The global operative assessment of laparoscopic skills (GOALS) has been accepted and validated as a training method for MIS [17]. With the use of GOALS, a trained expert assesses performance by watching a video of the task and provides a score on five elements: depth perception, bimanual dexterity, efficiency, tissue handling, and autonomy. Although the results are no longer biased because it is possible to perform a blind assessment, the evaluation still is subjective and requires a significant amount of time on the evaluator's part.

Another validated training method is the McGill Inanimate System for Training and Evaluation in Laparoscopic Surgery (MISTELS) [18], which has been incorporated into the fundamentals of laparoscopic surgery (FLS) curriculum as the manual skills component. The MISTELS method requires trainees to achieve proficiency for five basic tasks performed inside a physical simulator. A performance metric is calculated based mostly on the task completion time together with an evaluation of the final outcome of the task. Unfortunately, the evaluation of a final outcome still is a subjective measure that requires an expert evaluator, and it has been criticized because beginner trainees have not reached the automatous phase and should not be judged based on task completion time [17].

To address some limitations of the FLS curriculum, research has focused on the development of metrics that automatically assess performance during the entire task. The imperial college surgical assessment device (ICSAD) was developed for this purpose. It uses position sensors attached to the hands of the trainee and computes performance based on task completion time, number of movements, and total path length [19]. The ICSAD system still is limited in that it can evaluate only elements of performance related to motion and time and in that trainees must wear large sensors. However, a significant advantage of ICSAD is that it can be used in any training environment, including that of simulators.

Simulator-based training

Simulator-based training has been proposed as a means of developing surgical skills in MIS because the type of skills that need to be learned for MIS are easily trained with simulators [5]. Three different types of simulators are used: training boxes or physical simulators, virtual reality (VR) simulators, and augmented reality (AR) simulators. An excellent review of simulators is presented by Schreuder et al. [5].

Physical simulators consist of a training box that mimics the patient's body, with instruments entering through small openings (e.g., Simulab LapTrainer, Simulab Corporation, Seattle, WA [4]). These simulators have the advantage of being low cost, portable, adaptable, simple [4], and capable of providing realistic haptic feedback [20]. The main limitation of physical simulators is that they do not provide a measure of performance other than task completion time.

Virtual reality simulators are those that use a computer program to create a model of the surgical environment and the instruments. These types of simulators address the problem of lack of feedback by computing performance metrics based on movement of the instruments or the trainee's hands and their interactions with the virtual environment [21]. However, these simulators usually are costly and lack realistic haptic feedback [5].

Augmented reality simulators solve many of the aforementioned limitations by combining real environments with realistic haptic feedback and software programs that can enhance the surgical view, track instrument motion, provide performance metrics, and track trainee progress (e.g., the ProMIS system, CAE, Saint-Laurent, Quebec).

In general, for simulators to be effective, they must provide feedback during learning, allow trainees to repeat each skill, adapt progressively to more difficult tasks, provide individualized learning in a controlled environment, and provide well-defined outcomes [13]. However, this requires the availability of performance metrics that truly reflect performance.

Background on performance metrics

A significant amount of work has focused on finding performance metrics, which are required to determine the level of experience that surgeons and trainees demonstrate when performing specific tasks. The most commonly used performance metrics are presented in the following sections.

Temporal

Measurement of task completion time is a common way of assessing performance. Most currently used simulators and

metrics use time in one way or another to measure skill. Task completion time has been used for skills assessment in many studies [3, 16, 19, 20, 22–24] and may provide an indication of trainee skill levels when combined with other performance metrics. Looking at the time between subtasks [14] gives a measure of hesitation.

Outcome metrics

Outcome or qualitative metrics are those that assess the final outcome of each task or the procedure as a whole. These metrics do not analyze how the procedure was performed but instead are concerned only with the end result. Examples of this type of metric include the number of errors [22, 25], the number of attempts required to achieve the desired outcome, the quality of the outcome [14], and the specific criteria defined for each specific task [15].

Although the implementation of outcome metrics requires less time commitment on the part of the evaluator than that of the metrics based on checklists or standard rating scales, the process is still time consuming and subjective. Furthermore, the type of assessment must be very specific to the task being performed.

Motion-based metrics

Motion-based metrics are the metrics most commonly used for objective measurement of performance. The two most commonly used motion-based metrics are the number of movements [15, 19, 23] and the distance traveled (path length) [3, 20, 23, 25, 26]. The path length may be computed as follows [20]:

$$P = \int_0^D \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt, \quad (1)$$

where D is the task duration and the variables in the brackets correspond to the first derivative of the motion in the three Cartesian directions: x , y , and z .

Analysis of surgical gestures using hidden Markov models [24, 27] and multivariate autoregressive models [28] has been implemented to measure skill levels throughout a task. However, creating these models requires each step of the procedure to be categorized during the analysis. Other metrics that can be computed from position data are presented in the following sections.

Velocity and speed

Many performance metrics currently used are based on velocity and speed values. Velocity often is computed as the first derivative of the motion profile, whereas speed considers only the magnitude of the velocity vector. The

metrics proposed in the literature include a normalized speed metric (computed as the mean speed divided by the maximum speed) [29], the mean speed [3], the peak speed [3, 16], the instantaneous speed vector [26], the three-dimensional instrument tip velocity [30], the number of changes in velocity over time, and the number of peaks in speed [29]. The movement arrest period ratio (MAPR), used in Rohrer et al. [29] as a measure of how often the speed is near zero (measurement of hesitation), is defined as the proportion of time that the movement speed exceeds a given percentage of the peak speed.

Acceleration

Another metric commonly used is acceleration, computed as the second derivative of the motion profile. The metrics based on acceleration include the number of accelerations and decelerations [25], the mean acceleration [3], and the maximum acceleration [3]. Another metric is the integral of the acceleration vector (IAV), which measures the energy expenditure and is defined by Cavallo et al. [3] as follows:

$$\text{IAV} = \int_0^D \sqrt{\left(\frac{d^2x}{dt^2}\right)^2 + \left(\frac{d^2y}{dt^2}\right)^2 + \left(\frac{d^2z}{dt^2}\right)^2} dt. \quad (2)$$

Jerk

The third derivative of the motion profile, known as jerk, used by several researchers as a measure of motor skills, usually is applied for assessing the progress of certain diseases such as neurodegenerative diseases, injuries to the jaw [26], or the effects of experiencing a stroke [29]. Jerk has been shown to discriminate between healthy patients and those with motor dysfunctions and can be used to identify progress during learning [26]. It was proposed as a means of assessing MIS skill development by Cotin et al. [20] and by Hwang et al. [31]. Unfortunately, the latter study had insufficient power and failed to determine whether the jerk exhibited by novices differed from that exhibited by experts.

A limitation of the jerk metric is that it is inversely dependent on the second power of task completion time. Hence, it is not completely independent of task duration.

Several different ways of normalizing jerk have been proposed by Hogan and Sternad [32], and they show that a dimensionless metric remains constant as the amplitude of the motion (A_m) and that the duration varies. Jerk is sensitive to increases in the number of peaks, the amplitude of the peaks, and the periods of arrest, providing a real measure of smoothness. Based on this, a three-dimensional jerk metric is presented by Cavallo et al. [3]:

$$\text{Jerk}_{\text{norm}} = \sqrt{\frac{D^5}{2A_m^2} \int_0^D \left(\left(\frac{d^3x}{dt^3}\right)^2 + \left(\frac{d^3y}{dt^3}\right)^2 + \left(\frac{d^3z}{dt^3}\right)^2 \right) dt}. \quad (3)$$

Care still must be taken when the jerk metric is used to assess performance. It is important to note that smoothness will be measured as high if large pauses exist between movements, which makes jerk counterintuitive as a performance measure [29]. Therefore if a novice predominantly uses the dominant hand, the motion profile will show higher jerk than for the other hand.

Force-based metrics

Similar to position data, force data can be analyzed in many ways. New instruments and devices that allow force information to be measured during training have initiated the development of performance metrics that reflect the ability to be gentle or to apply sufficient force when required. Very little work has been done on the use of force information for skills assessment and training in MIS, limited by the capability to measure force in real surgery.

Applied forces may be an important measure to consider in characterizing trainee skill level, but this approach is not straightforward in determining what distinguishes an expert from a novice because ideal applied forces are task dependent [14]. Some VR simulators have been developed with objective assessment metrics based on the maximum forces applied [14] or on grasping with excessive pressure [21]. A study by Hwang et al. [31] presented a laparoscopic grasper instrumented with a force/torque sensor and strain gauges on the handle of the instrument to measure the applied forces during real surgical procedures. Unfortunately, apart from its ability to measure only the forces applied from outside the patient's body, this study was underpowered, and no significant differences were found between the forces applied by novices and experts.

An interesting study by Tang et al. [33] showed that trainees found it difficult to be gentle with tissue, a phenomenon often called “the heavy hands” of the beginner. This study found that force-related errors (too much or too little) dominated 58 % of consequential errors and 31 % of inconsequential errors.

Some of the force-based metrics proposed in the literature for skills assessment in MIS include the average force [16] and the maximum or peak force. The latter value is affected by outliers, so care must be taken when the information is interpreted. This metric did not show a difference between experience levels in the study of Dubrowski et al. [16].

Combination of metrics

Because performance is not affected by one factor alone, combining different performance metrics might provide a better way of measuring outcome. To compare various performance metrics, the metrics need to be within the same range and ideally unitless.

Different ways of normalizing performance data have been proposed in the literature. The method presented in Cotin et al. [20] and Stylopoulos and Vosburgh [34] compares each individual parameter with those obtained from a group of experts. This method provides a way of generating a combined performance metric from individual metrics, and an analysis using this method is presented by Trejos et al. [35]. This method for combining metrics is limited by its use of data from the expert group as part of the equation for determining the overall metric, which significantly impairs the objectivity of the metric.

Materials and methods

The aforementioned review shows a very clear need for the development of performance metrics that are automatically computed based on motion or force data, that are objective and do not rely on the user's input for assessment, that provide a measure of the performance throughout the task and not only the final outcome, and that provide a measure of aspects important to consider during surgery such as safety and dexterity. The following section describes an experimental evaluation that aimed to identify new performance metrics that meet these requirements and to establish how well they correlate with trainee experience level.

Experimental setup and methods

The sensorized instrument-based minimally invasive surgery (SIMIS) system [36] used in these experiments is composed of two sensorized laparoscopic instruments capable of measuring tool–tissue interaction forces and the position of the instrument tip. The system includes customized software that allows the following data to be recorded for each instrument: grasping force, torsion about the instrument axis, Cartesian forces, and position data in all six degrees of freedom. It also can record a video of the entire trial synchronized with the force and position data.

Because a large number of subjects were needed to complete these trials, a more robust version of the sensorized instruments was used in these experiments [37]. The increase in robustness reduced the sensitivity for sensing axial forces. Because the instruments enter the training box through side ports, the maximum forces are applied perpendicular to the

instrument shaft and not in the axial direction. Due to the small magnitude of the axial forces and the increased noise level present in the instruments used, it was decided to consider only the forces acting perpendicular to the shaft (i.e., in the x and y directions). These two forces were combined into one force value for the analysis.

To perform the experiments, it was first necessary to develop an appropriate experimental setup. Many simple tasks can be performed easily in a minimally invasive manner without force feedback. These tasks were not adequate to achieve the objectives of this work because they are too simple.

A complex procedure that required completion of both technical and cognitive skills and that was composed of tasks shown to require some form of force information was developed. The procedure involved five tasks: palpation of tissue to locate a lesion or tumor, intracorporeal suturing and knot tying, and cutting near a critical anatomic feature.

The setup comprised foam and silicone of different compositions. A 1-cm cylinder made of silicone rubber (Sorta-Clear 18, Shore hardness 18A; Sculpture Supply, Etobicoke, ON, Canada) was embedded in a tissue phantom (from the Chamberlain Group, Great Barrington, MA) to mimic soft tissue with an embedded tumor. A replaceable top skin surface made of soft rubber (EcoFlex, Shore hardness OO-30; Sculpture Supply Canada) was used to hide the lump visually. A plastic frame, designed and built from ABS plastic, was used to attach the model to the laparoscopic box and hold it in place. The locations of the tumors were varied randomly, and the subjects were blinded to the locations. This setup allowed participants to perform a complex procedure composed of the five following tasks:

Task 1: Palpation The SIMIS instruments were used to palpate the tissue to locate the tumor. This task usually was completed when they could see a lump (Fig. 1A).

Task 2: Cutting The instrument in the dominant hand was replaced by a set of standard laparoscopic scissors, which were used to cut the thin skin covering the tumor (Fig. 1B).

Task 3: Tissue-handling The SIMIS instruments were used to remove the tumor (Fig. 1C).

Task 4: Suturing The instruments were used to drive a needle through the tissue (Fig. 1D).

Task 5: Knot-tying An intracorporeal surgeon's knot composed of one double knot and two single knots was tied (Fig. 1E, F).

Institutional review board approval was obtained from Western University before the trials began. A total of 30 subjects (7 women and 23 men) performed the complex procedure four times. All the subjects were right-handed. As described in Table 1, the experience of the subjects varied based on background, postgraduate year (PGY) level, and years of practice.

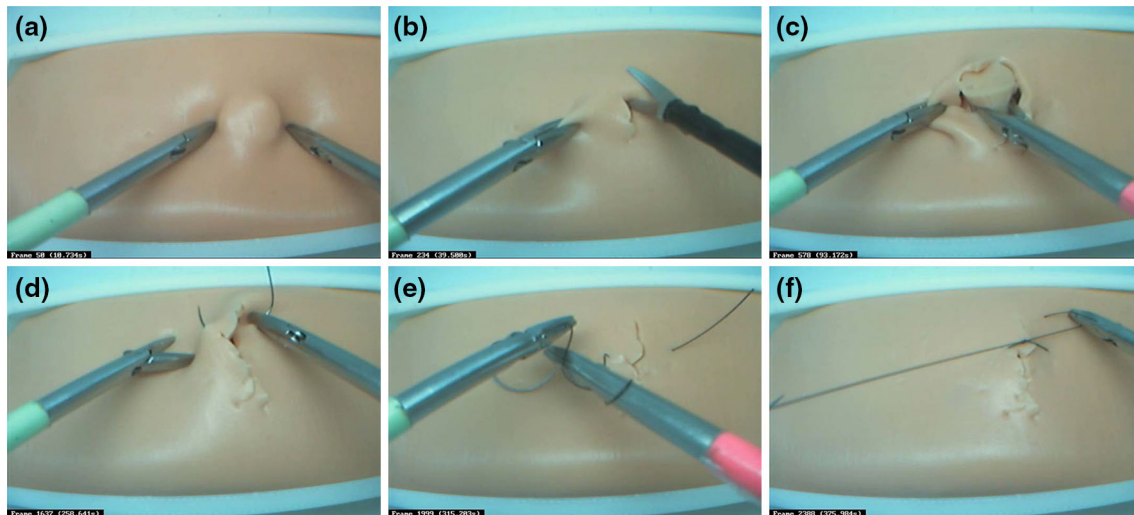


Fig. 1 Steps in a complex procedure composed of five tasks. **A** Palpate tissue to identify tumor location. **B** Cut top surface to expose the tumor. **C** Remove tumor. **D** Pass a suture. **E, F** Tie and tighten an intracorporeal surgeon's knot

Table 1 Categorization of subject experience levels

Basic experience level	Detailed experience level	Description
Novice ($n = 17$)	1 ($n = 6$)	No medical background (e.g., engineers)
	2 ($n = 6$)	Medical students
	3 ($n = 5$)	PGY 2–3 and surgeons with no MIS training
Expert ($n = 13$)	4 ($n = 2$)	PGY 4–5 with training
	5 ($n = 5$)	Fellows with training
	6 ($n = 6$)	Expert surgeons

PGY postgraduate year; MIS minimally invasive surgery

Data processing

The videos of all 120 trials were observed and analyzed as follows. The start and end times of each task were identified and recorded. Time frames were recorded for any events that were out of the ordinary (e.g., if the needle was dropped and no longer visible, the subject took a break, the instruments needed fixing, or the skin lifted off of the setup and needed to be replaced). The time frames corresponding to actions between the tasks were identified. This process was followed to reduce variability in the data because the subjects all were unique in their way of removing the instruments from the setup or dropping the tumor to the side. Tasks 4 and 5 had no dead time between them.

As described later, some of the metrics proposed in this article rely on computation of the first, second, and third derivatives of the force and position data.

Although the data were low-pass filtered at 10 Hz when recorded by the SIMIS software, this filtering was observed to be insufficient when the data needed to be differentiated (no additional filtering was implemented between derivatives). Therefore, a second-order Butterworth filter with a cutoff frequency of 1.25 Hz was applied to the data before computation of the first derivative. The MATLAB (The Mathworks, Inc., Natick, MA) `filtfilt` function was used, which filters data in the forward direction and then refilters the output in the reverse direction so that phase distortion is eliminated. A MATLAB script was run to separate the data into the different tasks, compute the total range of forces applied in each direction, and create individual plots for evaluation. The plots then were reviewed to identify any discrepancies in the data. The data then were processed to compute the performance metrics for the Cartesian force and the grasping force, as detailed in the following sections.

Position-based metrics

The position-based metrics proposed in the literature were computed as follows:

1. Total volume was computed by calculating the maximum and minimum positions in each direction and then multiplying the resulting three ranges of motion.
2. Interquartile volume was calculated by multiplying the interquartile ranges (IQRs) in each direction (using the `iqr` function from the MATLAB Statistics Toolbox).
3. Velocity was computed by calculating the first derivative of the motion profile for x , y , and z (using the

MATLAB diff function with a sampling time of 0.02 s). The three velocity components then were combined into a single speed magnitude through the Euclidean norm of each data point. The tip speed profile then was used to compute the following metrics:

- (a) The consistency of the speed was calculated as the standard deviation of the tip-speed profile.
 - (b) The number of peaks in the speed was calculated using the MATLAB findpeaks function to find the number of local peaks in the tip-speed profile.
 - (c) The peak speed was calculated as the maximum of the tip-speed profile.
 - (d) The average speed was calculated as the mean of the tip-speed profile.
 - (e) The MAPR was calculated as the proportion of time that the movement speed exceeded 25 % of the maximum speed.
 - (f) The path length was approximated by following Equation 1 and using the MATLAB trapz function.
4. Acceleration was computed by differentiating the velocity profiles in each direction and then combining the components using the Euclidean norm. This value was used to calculate the following metrics:
 - (a) The acceleration consistency was calculated as the standard deviation of the acceleration profile.
 - (b) The peak acceleration was calculated as the maximum of the acceleration profile.
 - (c) The average acceleration was calculated as the mean of the acceleration profile.
 - (d) The IAV was computed as the integral of the acceleration profile as defined in Equation 2.
 5. Normalized jerk was calculated by differentiating the acceleration profile in each direction and then combining the components using Equation 3.

Force-based metrics

The average and peak forces for both the Cartesian and the grasping forces were computed in this analysis. Other performance metrics that have not been used for skills assessment and training in MIS also were implemented as follows:

1. *Force range* The difference between the minimum and the maximum forces applied during a task is important because it encompasses the magnitude of the forces in both directions. The force range for the Cartesian and the grasping forces was computed.

2. *Interquartile range* This metric takes into account the 50 % of the data closest to the median so that outliers do not have an effect on the overall metric.
3. *Integral of the force* This value provides a measure of high forces and the amount of time that forces are high. The integrals of the grasping and the Cartesian force profiles also were approximated using the MATLAB trapz function with a sampling time of 0.002 s.
4. *Force derivatives* The first and second derivatives of the force could indicate consistency of force application. The vector of force derivatives was computed using the diff function. The derivative metric (dF_{metric}) then was calculated using the following equation:

$$dF_{\text{metric}} = \sqrt{\frac{D}{2F_{\text{iqr}}^2} \int_0^D \left(\frac{dF}{dt}\right)^2 dt}, \quad (4)$$

where F_{iqr} is the IQR of the force profile. Similarly, the vector of the second derivative of the force was computed by differentiating the first derivative, and the second derivative metric (d^2F_{metric}) was computed using the following equation:

$$d^2F_{\text{metric}} = \sqrt{\frac{D^3}{2F_{\text{iqr}}^2} \int_0^D \left(\frac{d^2F}{dt^2}\right)^2 dt}. \quad (5)$$

5. *Smoothness of the applied forces* The third derivative of the force provides a measure for the regularity and uniformity of the contact forces. The third derivative metric (d^3F_{metric}) was calculated using the following equation:

$$d^3F_{\text{metric}} = \sqrt{\frac{D^5}{2F_{\text{iqr}}^2} \int_0^D \left(\frac{d^3F}{dt^3}\right)^2 dt}. \quad (6)$$

Combined metrics

Combined force and position metrics were implemented to account for various important skills that need to be developed together. In the study of Beyer et al. [17], the GOALS score is computed as a combination of several different metrics. Following those same metrics and considering what could be measured with the SIMIS system, the following metrics were considered to be important:

1. *Depth perception* As a measure of depth perception, the GOALS score looks at overshooting targets. This can be related to motion smoothness (i.e., the jerk metric).
2. *Bimanual dexterity* Because the MAPR calculates the percentage of time that the instrument is being used, a measure of bimanual dexterity was developed by subtracting the MAPR value for the nondominant hand from the MAPR value for the dominant hand.

3. *Efficiency* As a measure of efficiency, the total volume used for each task and the number of peaks in speed were considered important.
4. *Tissue handling* This is a measure of how roughly the tissue is handled or whether any tissue damage occurs. The metrics considered for tissue handling included the integral and the derivative of the grasping and the Cartesian forces for both instruments.

Considering the metrics for the left and right hands, this resulted in a total of 15 metrics that needed to be combined. Because all the metrics have different units and different ranges, to combine them properly, the first step requires a normalization of each metric so that those metrics with higher values do not dominate over the remainder. To achieve this normalization, the following equation was implemented:

$$z_i = \begin{cases} \frac{P_i}{2P_i} & z_i < 5 \\ 5 & z_i \geq 5 \end{cases}, \quad (7)$$

where z_i is the normalized version of the i th metric, P_i is the value obtained by each trainee for that particular metric, and P_i is the trimmed mean of the data set (i.e., the mean of the values without the top three maximum and minimum values). This equation ensures that the closer the value is to 1, the closer it is to the mean of the entire group without consideration of the outliers. Furthermore, each metric is capped at 5 to ensure that the outlier data do not dominate. A total metric for each trainee then is computed as follows:

$$z = \sum_{i=1}^N \alpha_i z_i, \quad (8)$$

where N is the total number of metrics being combined and α_i represents the scaling coefficients that may be used to balance the influence of each parameter. By adjusting the value of α_i , it is possible to modify the weight or importance of each individual metric in the total combined metric. These equations allow the metrics to be combined such that one metric does not dominate over the others and such that they can be adapted to the task being performed.

To determine the scaling factors, an optimized scaling vector was calculated with the goal of maximizing Spearman's rho correlation with the experience level. The MATLAB `fmincon` function was used to find optimal parameters constrained between a lower and an upper bound (set at 0 and 1, respectively). It was used to find the set of scaling values that generated the minimum correlation (maximum negative correlation) between the metric and the experience level for each of the tasks.

Data analysis

The Statistical Package for the Social Sciences, version 19 (SPSS, Chicago, IL, USA) was used to perform statistical analysis of the data. An initial analysis was performed by comparing the metrics for the expert and the novices using an analysis of variance. The results, presented in [37], show that most metrics were able to show a statistically significant difference at the novice versus expert levels. The results presented in this article are aimed at measuring the correlation between the metrics and the six detailed levels of experience (Table 1). To measure this dependency, Spearman's rho correlation was computed to determine how well each of the different metrics correlated with the six experience levels. The results are presented in the following sections.

Results

The first analysis of the results showed that one of the experienced subjects created outlier data during the palpation task (task 1). This subject had difficulty locating the tumor at every try and needed to make as many as five incisions to locate the tumor. Because this was a statistically significant outlier confirmed using DesignExpert (based on the externally studentized residuals), it was removed from the data for the analysis of task 1 only.

The results of the Spearman's rho correlations between the metrics evaluated and the experience level are shown in Table 2. The average task completion time results also are presented in Fig. 2. The following sections highlight the results in more detail.

Time

Time showed a significant correlation with experience level in all tasks, decreasing as experience increased. The correlations were weak for the simpler tasks (-0.242 to -0.336 for tasks 1 to 3; $p < 0.05$) and became stronger as the task complexity increased (-0.437 for task 4 and -0.769 for task 5; $p < 0.05$). As shown in Fig. 2, a consistently decreasing trend could not be observed for any of the tasks. The suturing and knot-tying tasks (tasks 4 and 5) showed a plateau after experience level 4, which is the point at which students are considered trained in basic MIS tasks (Table 1).

Position

Not all of the position metrics showed significant correlations. Table 2 shows that the number of peaks in speed and

Table 2 Spearman's rho correlation between the six levels of experience and each metric evaluated^a

Metric	Task 1: Palpation (LH, RH)	Task 2: Cutting (LH only)	Task 3: Handling (LH, RH)	Task 4: Suturing (LH, RH)	Task 5: Tying (LH, RH)
Task completion time	-0.336	-0.242	-0.297	-0.437	-0.769
Total volume	-0.182, -0.197	0.020	-0.054, -0.122	-0.331, -0.223	-0.541, -0.528
Speed consistency	-0.220, -0.202	-0.114	-0.132, -0.116	-0.250, -0.177	-0.465, -0.357
Speed peaks	-0.338, -0.345	-0.237,	-0.330, -0.302	-0.445, -0.434	-0.767, -0.772
Maximum velocity	-0.062, -0.058	0.121	0.010, 0.150	-0.206, 0.006	-0.230, 0.024
Mean velocity	0.175, 0.291	0.117	0.171, 0.260	0.207, 0.546	0.270, 0.562
MAPR	0.151, 0.211	-0.092	0.095, -0.018	0.334, 0.262	0.481, 0.436
Path length	-0.305, -0.276	-0.154	-0.188, -0.195	-0.348, -0.316	-0.725, -0.680
Acceleration consistency	-0.235, -0.216	-0.138	-0.148, -0.127	-0.233, -0.173	-0.465, -0.340
Maximum acceleration	-0.102, -0.075	0.120	-0.005, 0.104	-0.142, -0.027	-0.196, 0.079
Mean acceleration	0.096, 0.255	0.204	0.151, 0.253	0.214, 0.555	0.252, 0.525
IAV	-0.319, -0.283	-0.155	-0.199, -0.180	-0.342, -0.316	-0.725, -0.671
Jerk	-0.341, -0.333	-0.243	-0.296, -0.268	-0.409, -0.406	-0.750, -0.736
IQR volume	-0.153, -0.065	0.156	0.081, -0.031	-0.319, -0.149	-0.285, -0.269
Grasp mean	-0.251, -0.038	0.039	-0.034, -0.112	-0.352, -0.290	-0.238, -0.191
Grasp maximum	-0.336, -0.107	0.024	0.034, -0.322	-0.388, -0.349	-0.369, -0.337
Grasp IQR	-0.181, 0.096	0.106	0.029, -0.045	-0.164, -0.234	-0.142, -0.179
Grasp integral	-0.320, -0.190	-0.082	-0.178, -0.386	-0.520, -0.523	-0.678, -0.666
Grasp first derivative	-0.386, -0.418	-0.053	-0.130, -0.363	-0.393, -0.524	-0.695, -0.721
Grasp second derivative	-0.330, -0.374	-0.139	-0.193, -0.384	-0.411, -0.514	-0.748, -0.769
Grasp third derivative	-0.306, -0.352	-0.179	-0.227, -0.357	-0.408, -0.505	-0.763, -0.782
Cartesian forces mean	-0.137, -0.142	-0.257	-0.109, -0.054	-0.159, -0.418	-0.406, -0.397
Cartesian forces maximum	-0.284, -0.356	-0.080	-0.059, -0.170	-0.304, -0.439	-0.371, -0.402
Cartesian forces IQR	0.044, 0.025	-0.029	0.011, -0.050	-0.109, -0.278	-0.488, -0.378
Cartesian forces integral	-0.293, -0.313	-0.376	-0.311, -0.339	-0.454, -0.567	-0.755, -0.780
Cartesian forces first derivative	-0.355, -0.381	-0.129	-0.203, -0.256	-0.459, -0.437	-0.729, -0.652
Cartesian forces second derivative	-0.322, -0.315	-0.228	-0.260, -0.290	-0.470, -0.460	-0.769, -0.729
Cartesian forces third derivative	-0.303, -0.291	-0.247	-0.283, -0.302	-0.465, -0.450	-0.776, -0.748

Correlations in bold type are significant with $p < 0.5$

MAPR movement arrest period ratio; IQR interquartile range; LH left hand; RH right hand

the normalized jerk exhibited significant correlations for all of the tasks with a p value lower than 0.05. Correlations between experience level and speed peaks were weak for the simpler tasks (-0.237 to -0.345 for tasks 1 to 3), intermediate for task 4 (-0.445 for the left hand and -0.434 for the right hand), and strong for task 5 (-0.767 for the left hand and -0.772 for the right hand). Similarly, the correlations between the experience level and the jerk were weak for the simpler tasks (-0.243 to -0.341 for tasks 1 to 3), intermediate for task 4 (-0.409 for the left hand and -0.406 for the right hand), and strong for task 5 (-0.750 for the left hand and -0.736 for the right hand).

A closer look at the speed peaks showed that they were directly coupled with the task completion time, as evidenced by the same shaped graphs for all the tasks. The average normalized jerk provided a better measure of

performance, as shown in Fig. 3. It can be observed in this graph that most of the tasks had a decrease in the jerk as the experience level increased. However, this decrease also tended to plateau after experience level 4. Some of the correlations found with the position metrics were slightly stronger than the correlations found with task completion time.

Force

Compared with time and position, stronger correlations were observed in some of the force-based metrics, as shown in Table 2. The correlations between the force-based metrics and the experience levels for tasks 1–3 were weak, with significant correlations of -0.18 to -0.42 . The correlations were significant for most of the metrics during

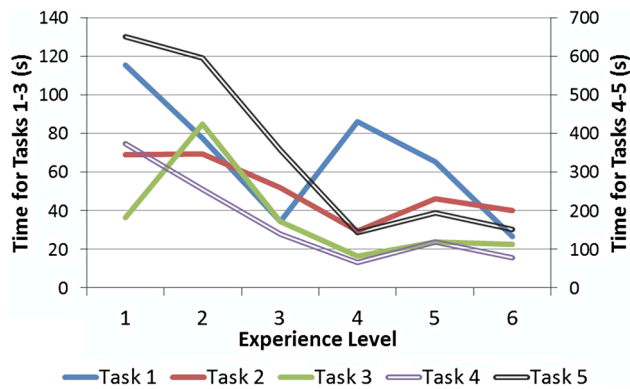


Fig. 2 Average task completion time for the five tasks according to the level of experience

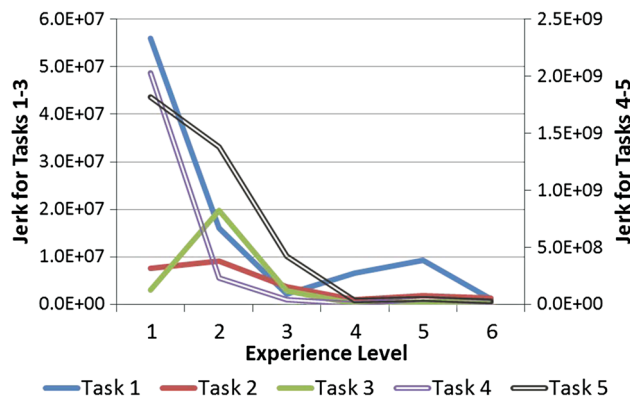


Fig. 3 Average normalized jerk as a function of experience level for all five tasks

tasks 4 and 5, ranging from -0.23 to -0.57 for task 4 and from -0.20 to -0.78 for task 5, with the strongest correlations present in the second and third derivatives of the applied forces during task 5 (-0.73 to -0.78). More important, however, is the fact that some of the metrics showed a trend toward a consistently decreasing slope for the palpation task, as well as during suturing and knot tying (tasks 1, 4, and 5). Some examples of these metrics are shown in Fig. 4.

Combined metrics

The combined metric was implemented and evaluated. Figure 5 shows a comparison between the correlations found with the task completion time, the peaks in speed, jerk, the integral of the force, and the optimized combined metric. This combined metric combined the jerk metric, the difference in the MAPR value between the two hands, the total volume, the number of peaks in speed, and the integrals and derivatives of the grasping and Cartesian forces. The values of this combined metric were -0.50 for task 1, -0.37 for task 2, -0.43 for task 3, -0.61 for task 4, and

-0.85 for task 5. The scaling factors were determined through an optimization strategy that aimed to find the strongest correlations with experience level. This figure shows that the force-based metrics and the combined metric exhibited stronger correlations with experience level than the task completion time or the position-based metrics alone.

The results of the optimization for the combined metric are shown in Table 3. Interestingly, the metrics that dominated the combined metric were force-based, with the exception of volume, which was important during the suturing task (task 5).

Discussion

Spearman's rho correlation was chosen in the aforementioned analysis as a way to quantify the relationship between each metric and subject experience levels. The study presented in this article was aimed at observing the correlation between the proposed metrics and the known levels of experience of the subjects.

Interestingly, for experience level 4 (subjects at the PGY 4–5 levels, see Table 1), the task completion times for tasks 2–4 were shorter than those for all the other groups. Other studies have shown similar results, for example, the study by Stefanidis et al. [38]. This can be explained by the fact that these trainees had recently completed their MIS training, in which time was the main measure of performance. In fact, Stefanidis et al. [38] showed that a clear decline in performance occurs after training (posttest evaluations) and an even further decline in retention tests when performance is assessed using the FLS metrics (which are mainly time based).

Because time is easy to measure, task completion time often is used as a performance metric. It is clear that for any kind of activity we perform, the more experience we have, the faster we can perform the task. However, care must be taken when time is used as a performance metric for several reasons:

1. Performing a task quickly means that the trainee has reached the automatic phase [17] but does not mean that the task is being performed correctly.
2. A clear trade-off exists between speed and accuracy; hence, performing a task faster is not necessarily better.
3. Everyone is different, and what is fast for one person might not be fast for another. Time is not a measure of ability [12], and it is important for surgeons to work at their own pace, especially near critical areas.
4. Depending on the specialty, doing things too fast could be a detriment to the overall outcome. This is

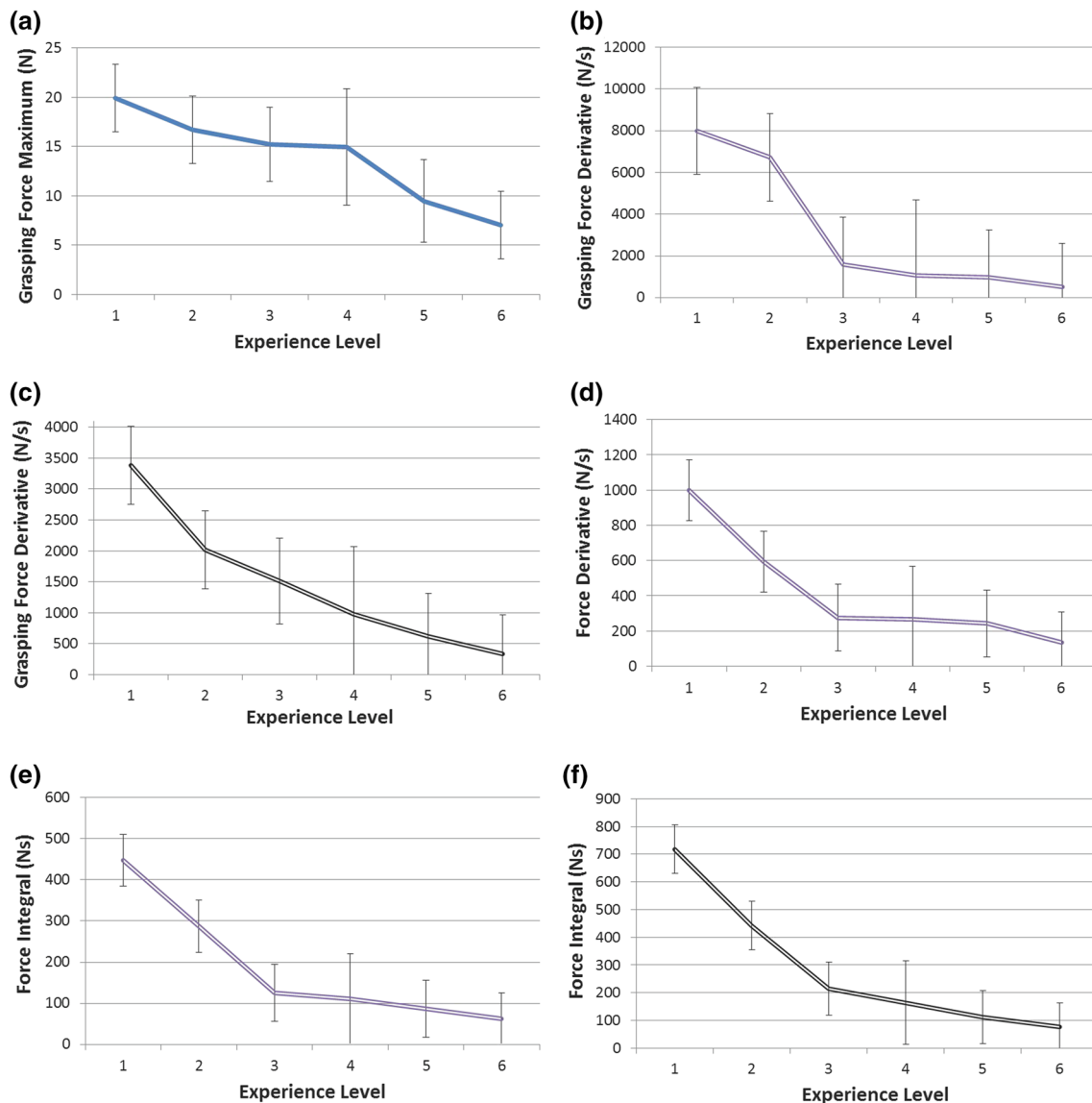


Fig. 4 Sample graphs of average force-based metrics across all subjects. **A** Maximum grasping force for task 1. **B** Derivative of the grasping force for task 4. **C** Derivative of the grasping force for task

5. **D** Derivative of the Cartesian force for task 4. **E** Integral of the Cartesian force for task 4. **F** Integral of the Cartesian force for task 5. *Error bars* correspond to \pm one standard deviation

especially true for thoracic surgeons who work close to critical anatomic features.

5. Training for time teaches trainees to focus on doing a task fast, and they may become aggressive to achieve the time requirements.
6. An overall time metric might be influenced by other aspects of the training scenario, for example, distracting factors or other differences between the practice scenario and the assessment scenario.

Nevertheless, task completion time may be useful as a measure of trainee skill level when combined with other metrics. The results of the position- and force-based metrics show interesting trends for qualifying experience during a complex procedure composed of five tasks. Some

of the position-based metrics and most of the proposed force-based metrics showed significant correlations with the six levels of experience ($p < 0.05$). As expected, the correlations found for the simpler tasks (tasks 1–3) are weak, whereas those found for the complex tasks (tasks 4 and 5) are the strongest.

The strongest correlations with the position-based metrics were found with the speed peaks and jerk metrics (Table 2). However, the correlations found with these metrics and experience level were not much stronger than those found with time except for the speed peaks during the tissue-handling task (task 3). Table 2 also shows that a few of the force-based metrics exhibited greater correlations with experience level than those found with time and

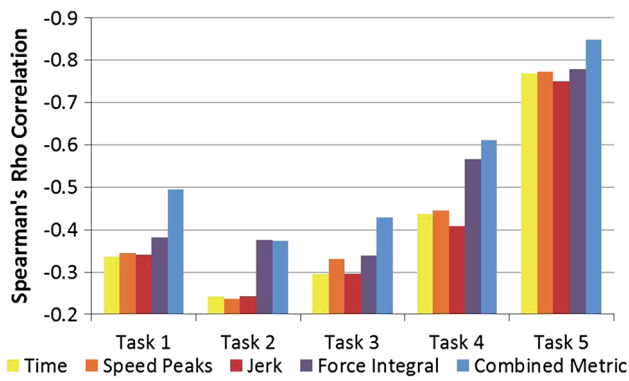


Fig. 5 Comparison of the best possible Spearman's rho correlations between the six levels of experience and several different metrics

Table 3 Scaling factors resulting from the optimization of the combined metric

Jerk RH	0.07	0.01	0.03	0.02	0.21
Jerk LH	0.05	0.00	0.02	0.02	0.19
MAPR	0.01	0.00	0.03	0.01	0.02
Volume LH	0.01	0.00	0.01	0.01	0.32
Volume RH	0.01	0.00	0.01	0.03	0.11
Peaks LH	0.04	0.01	0.03	0.02	0.20
Peaks RH	0.04	0.00	0.02	0.02	0.24
Force int. LH	0.18	0.98	0.02	0.14	0.11
Force int. RH	0.03	0.00	0.45	0.78	0.59
Force der. LH	0.29	0.00	0.01	0.42	0.09
Force der. RH	0.32	0.00	0.04	0.04	0.07
Grasp int. LH	0.38	0.00	0.01	0.74	0.40
Grasp int. RH	0.01	0.00	0.66	0.11	0.15
Grasp der. LH	0.36	0.00	0.01	0.02	0.46
Grasp der. RH	0.57	0.00	0.67	0.93	0.38

LH left hand; *RH* right hand; *MAPR* movement arrest period ratio; *int.* integral; *der.* derivative

motion. The strongest correlations were observed with the integral and the derivatives of the forces.

The results of the aforementioned experiments show that force-based metrics were able to provide stronger correlations with experience than those found with task completion time or position-based metrics. The relationships obtained with force showed consistently decreasing trends. With more subjects and increased power in the data, force-based metrics may be able to distinguish better between sublevels within the expert category because the trends show continuously decreasing values at the different levels. In other words, when trainees are considered trained in basic skills, time- and position-based metrics provide a measure of proficiency similar to that achieved by expert surgeons with many years of experience. However, some force-based metrics may be able to distinguish between those different levels.

With regard to the combined metrics, it should be noted that the original GOALS scores are intended to be measured by a trained expert based on visual observation. It is a very subjective metric that most likely would vary between evaluators. The five elements are assessed on a scale of 1–5 and then added up for a maximum score of 25 points. The analysis presented in this article aimed to provide some objectivity for the elements assessed by GOALS. Although these elements could be measured in other ways, this analysis proposes metrics that may be used to represent the GOALS score in a more quantitative manner. No current proof exists to show that the metrics selected are in fact directly related to the GOALS scores. These metrics were selected based on intuition considering the available metrics and the performance elements that needed to be measured.

Depth perception is something that cannot be directly measured without information on the target location. The jerk metric, as a measure of smoothness, was selected to measure depth perception because overshooting targets would cause the movement of the instruments to be less smooth. Other reasons not directly related to depth perception exist to explain why jerk would be high. The efficiency metrics were selected to represent efficiency in the use of space (the total volume used) and efficiency in the movement itself (number of peaks in speed). Instead of combining all the available metrics, the metrics presented earlier were proposed as a starting point in the quantification of those metrics considered to be important.

An interesting analysis results from observing the optimized scaling values presented in Table 3. From these parameters, we can identify which metrics are more affected by trainee experience levels. As the results show, the metrics that appear to be the most important for the combined force–position metric include the Cartesian force integral (tasks 2–5), the grasping force integral, and the derivative (tasks 1, 3–5), as well as the total volume (task 5). This combined metric provided stronger correlations with experience than any of the other single metrics. However, their implementation is more difficult because they depend on systems that can measure instrument motion and applied forces in the different degrees of freedom (i.e., grasping and Cartesian directions).

The value of computing a combined metric is that by adjusting the scaling coefficients, it may be possible to obtain metrics tailored to the task or procedure needing to be assessed. For example, they can serve to penalize severely for lack of accuracy in certain times or to penalize for lack of efficiency in others.

Conclusions

This study evaluated the effect of experience level on performance when a complex procedure composed of five

tasks was carried out, with the goal of identifying new performance metrics. Novel force-based metrics and metrics that combine force and position metrics were presented. These new metrics can be automatically computed, can provide a measure throughout the task, and can objectively measure aspects of performance that actually may have an effect on the outcome and safety of the procedure. The results show that experience level correlates well with force-based metrics. In particular, the integral and the derivative of the forces, or the metrics that combine force and position, provide the strongest correlations.

Future work in this area should evaluate the effect that training with the use of force-based metrics may have on the trainees' development and learning curve and should identify other possible combinations of metrics that include task completion time and the outcome of the procedure.

Acknowledgments The authors thank Dr. Michael Naish, Andrew Lyle, Abelardo Escoto, Chris Ward, Kevin Barker, Jerry Dafoe, and Dr. Ives Bureau for their assistance with several aspects of the work presented in this article and all the surgeons and trainees who willingly participated in the experiments. The support from Quanser, Inc., Markham, Ontario, Canada, is gratefully acknowledged. was supported by a grant from the Western Innovation Fund (R.V. Patel); by the Natural Sciences and Engineering Research Council (NSERC) of Canada under grants I2IPJ 363985-07 and RGPIN-1345 (R.V. Patel); by the Canada Research Chairs Program (R.V. Patel); and by infrastructure grants from the Canada Foundation for Innovation awarded to the London Health Sciences Centre (Canadian Surgical Technologies & Advanced Robotics [CSTAR]) and to The University of Western Ontario (UWO) (R.V. Patel). Financial support for Dr. Trejos was provided by an NSERC Alexander Graham Bell Canada Graduate Scholarship.

Disclosures Ana Luisa Trejos and Rajni V. Patel are inventors and have a patent application for the sensorized surgical instruments. The performance metrics presented in this article are not part of the patent application. They have no other conflicts of interest and no financial ties to disclose. Richard A. Malthaner and Christopher M. Schlachta have no conflicts of interest or financial ties to disclose.

References

1. Wulf G, Shea C, Lewthwaite R (2010) Motor skill learning and performance: a review of influential factors. *Med Educ* 44:75–84
2. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM (2004) Relationship between objective assessment of technical skills and subjective in-training evaluation in surgical residents. *J Am Coll Surg* 198:105–110
3. Cavallo F, Megali G, Sinigaglia S, Tonet O, Dario P (2006) A biomedical analysis of a surgeon's gesture in a laparoscopic virtual scenario. In: Westwood JD, Haluck RS, Hoffman HM, Mogel GT, Phillips R, Robb RA, Vosburgh KG (eds) *Medicine meets virtual reality. Studies in health technology and informatics*, vol 119. IOS Press, Long Beach, pp 79–84
4. Dayan AB, Ziv A, Berkenstadt H, Munz Y (2008) A simple, low-cost platform for basic laparoscopic skills training. *Surg Innov* 15:136–142
5. Schreuder HWR, Oei G, Maas M, Borleffs JCC, Schijven MP (2011) Implementation of simulation in surgical practice: minimally invasive surgery has taken the lead: the Dutch experience. *Med Teach* 33:105–115
6. Khan MS, Bann SD, Darzi A, Butler PEM (2003) Assessing surgical skill. *Plast Reconstr Surg* 112:1886–1889
7. Elliott D, Hansen S, Lyons J, Grierson LEM, Bennett SJ, Hayes SJ (2010) Goal-directed aiming: two components but multiple processes. *Psychol Bull* 136:1023–1044
8. Elliott D, Helsen WF, Chua R (2001) A century later: Woodworth's (1899) two-component model of goal-directed aiming. *Psychol Bull* 127:342–357
9. Elliott D, Grierson LEM, Hayes SJ, Lyons J, Elliott D (2011) Action representations in perception, motor control, and learning: implications for medical education. *Med Educ* 45:119–131
10. Choi Y, Qi F, Gordon J, Schweighofer N (2008) Performance-based adaptive schedules enhance motor learning. *J Mot Behav* 40:273–280
11. Reiley C, Lin H, Yuh D, Hager G (2011) Review of methods for objective surgical skill evaluation. *Surg Endosc* 25:356–366
12. Goff B, Mandel L, Lentz G, Vanblaricom A, Oelschlager A-MA, Lee D, Galakatos A, Davies M, Nielsen P (2005) Assessment of resident surgical skills: is testing feasible? *Am J Obstet Gynecol* 192:1331–1338
13. Lammers RL, Davenport M, Korley F, Griswold-Theodorson S, Fitch MT, Narang AT, Evans LV, Gross A, Rodriguez E, Dodge KL, Hamann CJ, Robey WC (2008) Teaching and assessing procedural skills using simulation: metrics and methodology. *Acad Emerg Med* 15:1079–1087
14. Moody L, Baber C, Arvanitis TN, Elliott M (2003) Objective metrics for the evaluation of simple surgical skills in real and virtual domains. *Presence Teleoper Virtual Environ* 12: 207–221
15. Datta V, Mandalia M, Mackay S, Chang A, Cheshire N, Darzi A (2002) Relationship between skill and outcome in the laboratory-based model. *Surgery* 131:318–323
16. Dubrowski A, Sidhu R, Park J, Carnahan H (2005) Quantification of motion characteristics and forces applied to tissues during suturing. *Am J Surg* 190:131–136
17. Beyer L, Troyer JD, Mancini J, Bladou F, Berdah SV, Karsenty G (2011) Impact of laparoscopy simulator training on the technical skills of future surgeons in the operating room: a prospective study. *Am J Surg* 202:265–272
18. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL (1998) Development of a model for training and evaluations of laparoscopic skills. *Am J Surg* 175:482–487
19. Bann S, Kwok K-F, Lo C-Y, Darzi A, Wong J (2003) Objective assessment of technical skills of surgical trainees in Hong Kong. *Br J Surg* 90:1294–1299
20. Cotin S, Stylopoulos N, Ottensmeyer M, Neumann P, Bardsley R, Dawson S (2005) Surgical training system for laparoscopic procedures. US Patent US2005/0 142 525 A1, 30 June
21. Oropesa I, Lamata P, Sanchez-Gonzalez P, Pagador J, Garcia M, Sanchez-Margallo F, Gomez E (2011) Virtual reality simulators for objective evaluation on laparoscopic surgery: current trends and benefits. In: Kim J (ed) *Virtual reality*. Intech Open Access Publisher, Rijeka, pp 349–374
22. Cao C, Zhou M, Jones D, Schwaizberg S (2007) Can surgeons think and operate with haptics at the same time? *J Gastrointest Surg* 11:1564–1569
23. Datta V, Mackay S, Mandalia M, Darzi A (2001) The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg* 193:479–485
24. Rosen J, Hannaford B (2001) Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills". *IEEE Trans Biomed Eng* 48:579–591

25. Boyle E, Al-Akash M, Gallagher AG, Traynor O, Hill ADK, Neary PC (2011) Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. *Postgrad Med J* 87:524–528
26. Takada K, Yashiro K, Takagi M (2006) Reliability and sensitivity of jerk-cost measurement for evaluating irregularity of chewing jaw movements. *Physiol Meas* 27:609–622
27. Megali G, Sinigaglia S, Tonet O, Dario P (2006) Modelling and evaluation of surgical performance using hidden Markov models. *IEEE Trans Biomed Eng* 53:1911–1919
28. Loukas CG, Georgiou E (2011) Multivariate autoregressive modeling of hand kinematics for laparoscopic skills assessment of surgical trainees. *IEEE Trans Biomed Eng* 58:3289–3297. <http://dblp.uni-trier.de/db/journals/tbe/tbe58.html#LoukasG11>
29. Rohrer B, Fasoli S, Krebs H, Hughes R, Volpe B, Frontera W, Stein J, Hogan N (2002) Movement smoothness changes during stroke recovery. *J Neurosci* 22:8297–8304
30. Cristancho S, Hodgson A, Panton O, Meneghetti A, Warnock G, Qayumi K (2009) Intraoperative monitoring of laparoscopic skill development based on quantitative measures. *Surg Endosc* 23:2181–2190
31. Hwang H, Lim J, Kinnaird C, Nagy A, Panton O, Hodgson A, Qayumi K (2006) Correlating motor performance with surgical error in laparoscopic cholecystectomy. *Surg Endosc* 20:651–655
32. Hogan N, Sternad D (2009) Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *J Mot Behav* 41:529–534
33. Tang B, Hanna G, Cuschieri A (2005) Analysis of errors enacted by surgical trainees during skills training courses. *Surgery* 138:14–20
34. Stylopoulos N, Vosburgh KG (2007) Assessing technical skill in surgery and endoscopy: a set of metrics and an algorithm (C-PASS) to assess skills in surgical and endoscopic procedures. *Surg Innov* 14:113–121
35. Trejos AL, Patel RV, Naish MD, Malthaner RA, Schlachta CM (2013) The application of force sensing to skills assessment in minimally invasive surgery. *IEEE International Conference on Robotics and Automation, Karlsruhe*, 6–10 May
36. Trejos AL, Patel RV, Naish MD, Schlachta CM (2009) A sensorized instrument for skills assessment and training in minimally invasive surgery. *ASME J Med Devices* 3:041 002-1–041 002-12
37. Trejos AL (2012) A sensorized instrument for minimally invasive surgery for the measurement of forces during training and surgery: development and applications. PhD Dissertation, Department of Electrical and Computer Engineering, The University of Western Ontario, Canada
38. Stefanidis D, Hope WW, Korndorffer JR, Markley S, Scott DJ (2010) Initial laparoscopic basic skills training shortens the learning curve of laparoscopic suturing and is cost-effective. *J Am Coll Surg* 210:436–440