

Contribution to the objective assessment of technical skills for surgery students: An accelerometer based approach

Giovanny Arbelaez Garces, David Joseph, Mauricio Camargo, Nguyen Tran, Laure Morel

▶ **To cite this version:**

Giovanny Arbelaez Garces, David Joseph, Mauricio Camargo, Nguyen Tran, Laure Morel. Contribution to the objective assessment of technical skills for surgery students: An accelerometer based approach. International Journal of Industrial Ergonomics, 2018, 64, pp.79 - 88. 10.1016/j.ergon.2017.12.002 . hal-01812007

HAL Id: hal-01812007
https://hal.univ-lorraine.fr/hal-01812007

Submitted on 29 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contribution to the objective assessment of technical skills for surgery students: an accelerometer-based approach

Authors: Giovanny Arbelaez-Garces[1*], David Joseph[2,3,4*], Mauricio Camargo[1], Nguyen Tran[4], Laure Morel[1]

[1] Université de Lorraine, ERPI, EA 3767, F-54000 Nancy, France
[2] Department of periodontology, Faculty of Dentistry, University of Lorraine, 54000 Nancy, France
[3] UMR-S1116, University of Lorraine, 54000 Nancy, France
[4] School of Surgery Nancy-Lorraine, University of Lorraine, 54505 Vandoeuvre-les-Nancy, France

*Co-first authorship statement: these authors contributed equally to this work.

Abstract:

Aim of study: Gestural skills assessment in odontology is a highly complex task. Although mandatory for certification, proficiency assessment is still required for parameters that are more objective. Our aim was to assess whether accelerometer sensors might contribute to distinguishing efficiency in therapeutic gesture from experts and novices in the realization of surgical sutures.

Material and method: 3 groups of participants: novices (n = 8), intermediaries (n = 14) and experts (n = 7) were enrolled in the study. They had to perform different types of odontological suture. We used 2 wireless wrist sensors to measure objectively the acceleration of both hands and to deduce the economy and the fluidity of the movement. A video recording was also used to complete the assessment.

Results: The time and motion criteria were documented to significantly correlate with the experience of the participant. The total time factor has a significant effect with the experience level ($p = 0.006$). The fluidity of the movement has also been shown to vary significantly between groups.

Conclusion: The use of accelerometers coupled with image analysis could make it possible to envisage in the long run an objective evaluation for this type of surgical gesture.

Keywords: Objective assessment, basic skills, suturing, dental pedagogic strategy, accelerometers

1. Introduction

The use of simulators and inanimate models for surgical training has slowly spread in training institutions around the world during the last ten years. Simulators allow trainers the freedom to assess and manage the complexity of the clinical scenario for students without risking patients' safety (Acton, 2015). The emergence of these technologies has not only revolutionized the training of surgical skills, but the assessment of surgeons in these techniques has become an important concern (Pellegrini, 2012). Moreover, being able to objectively assess surgeons' technical skills has become an important research topic and a principal concern for surgical training.

Most screen-based simulators allow direct assessment of gesture and enable immediate feedback. However, for other types of technologies such as inanimate models, an expert surgeon performs the assessment. Traditional skill assessment is subjective to the evaluator and can be influenced by several factors. To limit this problem, different strategies have been implemented, such as OSCE (objective structured clinical evaluation) (Scott, Evans, Drummond, Mossey, & Stirrups, 2001), analysis, videos, and more recently haptic simulation (Joseph et al., 2014). This highlights the need for a more objective assessment (Shaharan, 2014). The aim of this paper is thus to propose a method for objective surgical skills assessment using motion sensors.

(Datta, Mackay, Mandalia, & Darzi, 2001) proposed using motion-tracking analysis to objectively measure surgical skills. Their study suggested that hand motion analysis can be an effective objective

measure of dexterity. However, their study focused only on quantity of hand movements and time to complete the task. (Sánchez et al., 2014) proposed a method of evaluating laparoscopic skills using accelerometers. However, their work only focused on economy of movement criteria. In order to assess the complexity of the surgical skills put into practice during a procedure, several criteria need to be assessed, some of which might not be covered by motion sensors. In this paper, the complexity of surgical skills assessment is covered by completing the motion sensor data with criteria from an OSCE (Hanson, Mossey, & Macluskey, 2010; Niitsu et al., 2013). To validate the use of motion sensors (accelerometers) as a new objective evaluation strategy, we put in place a preliminary prospective study at the Nancy-Lorraine Surgery School. The study allowed us to check whether there is a relationship between the measurements made by motion sensors and a subjective evaluation when performing a basic surgical procedure in dentistry (sutures). A mixed method integrating objective and subjective criteria was then proposed and validated. The results obtained were then compared with the traditional OSCE based evaluation method.

The remainder of the paper is structured as follows: Section 2 presents an overview of research related to surgical skill assessment and particularly on the usage of accelerometers. Section 3 describes materials and methods used in the study. Sections 3.2 present the results, which are then discussed in Section 4. Section 5 presents the conclusions of the paper.

2. Related research

2.1. Surgical training and skills assessment

According to Fitts and Posner's theory of motor skill acquisition, it is divided into three stages (Fitts & Posner, 1967). The first stage is cognition; in this stage the learner is shown the basic elements of a surgical intervention, making a knot for example. The learner must understand the mechanics related to the skill, the way to move the hands, take the fill, etc. With practice, the learner goes through the integration stage; he becomes able to execute the task in a fluid manner with few interruptions. Then, in the automation stage, the activity is not only fluid but is carried out with almost no errors. When applied to surgical training, the first two steps of this model can be done without prior contact with patients through the acquisition of the basic skills of the surgical technique.

Despite technological advances, surgical education has remained almost unaltered for over a century. The "see one, do one, teach one" model, which until a few years ago was offering good results, is limited by the availability of expert surgical physicians, by the diversity of procedures, the amount of patients available and, more importantly, by the need to maintain patient safety (Morris, 2005). There are many surgical training devices that can be used in the learning process: inanimate models, virtual reality, synthetic and cadaveric animal models, as well as real patients for education of critical incidents (Acton, 2015; R. K. Reznick & MacRae, 2006). Some of these models are very expensive and some have limited availability. Using animals is difficult because of ethical concerns, high cost, and the need for special equipment. On the other hand, inanimate synthetic models are portable, reliable, and a little less expensive than earlier varieties.

Inanimate models are, however, at a disadvantage compared to virtual simulators because the latter allow us to directly assess the technical skill of the student without the intervention of an expert trainer. In order to objectively evaluate skills, performance verification lists or OSATS (Objective Structured Assessment of Technical Skills) are commonly used (R. Reznick, Regehr, MacRae, Martin, & McCulloch, 1997). However, the use of OSATS requires the implication of an experienced practitioner as judge. Technological tools could be used to assess some skill parameters and can become complementary to the objectivity of the evaluator (R. K. Reznick & MacRae, 2006).

One of the challenges in skill assessment is to choose which parameters to measure (van Hove, Tuijthof, Verdaasdonk, Stassen, & Dankelman, 2010). Hand movement analysis allows measurement

of motricity and dexterity (Mason, Ansell, Warren, & Torkington, 2013). In surgical skill assessment, this analysis can easily be performed with a motor sensor or accelerometer (D'Angelo et al., 2015). Three parameters have mainly been used in studies concerning surgical skills: time, economy of movement, and path length. Table 1 summarizes the main studies concerning movement analysis in surgical skills assessment and the parameters that were used.

Table 1. Movement analysis metrics used for surgical skills assessment.

Reference	Conceptual validation		
	Time	**Economy of movement**	**Path Length**
(Datta et al., 2001)	Yes	No	Yes
(Aggarwal et al., 2007)	Yes	Yes	Yes
(Datta, Chang, Mackay, & Darzi, 2002)	Yes	Yes	Non-treated
(Brydges, Classen, Larmer, Xeroulis, & Dubrowski, 2006)	Yes	Yes	Non-treated
(Brydges, Sidhu, Park, & Dubrowski, 2007)	Yes	Yes	Non-treated
(Smith, Torkington, Brown, Taffinder, & Darzi, 2002)	Yes	Yes	Yes
(Moorthy et al., 2004)	Yes	Non-treated	Yes
(S. D. Bann, Khan, & Darzi, 2003)	Yes	Yes	Non-treated
(Brydges et al., 2006)	Yes	Yes	Non-treated

2.2. Assessment metrics

The metrics for assessing surgical skills performance can be divided into two categories: efficiency metrics and quality metrics (Fried & Feldman, 2008). Efficiency metrics are measurable physical parameters, requiring the use of motion devices to be acquired. They are objective, reproducible, and not prone to misinterpretation. An efficiency metric should not only provide information about the intervention, but also help to identify key success and failure factors, in addition to being able to distinguish an expert from a novice (Oropesa et al., 2011). Quality metrics concern the definition and execution of a task; they are parameters that cannot be quantified – for example, knowledge of the procedure, final outcome of the intervention, knowledge of instrument handling, etc. Table 2 and Table 3 summarize the most commonly used metrics to discern the surgeon's experience.

Table 2. Efficiency metrics for surgical skills assessment.

Efficiency metrics	***Description***	***Units***
Time	The time needed to complete an intervention. It's commonly related to surgical skills. The greater the experience in the intervention, the less time it takes to complete.	*min, sec*
Path Length	The cumulative distance based on tool movement during the intervention. It is also commonly related to the surgeon's precision.	m, cm, mm

Jerk	These are sudden movements resulting in changes in the acceleration.	$\frac{m}{sec^3}$
Speed	The speed of change of position of the instruments per second.	$\frac{m}{sec}$
Economy of movement (MOE)	Execution of the intervention with a limited number of movements. It is defined as the total number of hand movements per second (S. Bann, Darzi, Munz, Kumar, & Moorthy, 2004).	$\frac{movements}{sec}$

Table 3. Quality metrics for surgical skill assessment.

Quality metrics	***Description***
Results	Analysis of the intervention outcome
Errors	Number of errors during the intervention
Instrument handling	Proper knowledge and use of the instruments during the procedure
Knowledge of specific procedure	Knowledge of the entire procedure

It appears that using a single type of metric to evaluate the surgeon's skills is not enough, but the assessment must be a combination of the two types (Fried & Feldman, 2008). Using only one of the metrics mentioned above may be misleading; for example, in the case of a suture, the time required to complete the task is irrelevant if the point does not approach the tissue correctly or the node is poorly executed. For these reasons it is advised to use the two types of metrics (Ponton-Carss, Kortbeek, & Ma, 2016).

As seen from the previous tables, efficiency metrics require the use of different technologies to be evaluated. Accelerometers and motion analysis have been used in the evaluation of the next section, which discusses some of the existing works.

2.2.1. Motion analysis usage

The use of motion-tracking analysis to assess surgical skills has already been proposed by (Datta et al., 2002, 2001) using electromagnetic devices and video analysis of hand movement. Their study suggested that hand motion analysis can be an effective objective measure of dexterity. However, their study focused only on quantity of hand movements and time to complete the task. (Sánchez et al., 2014), on the other hand, proposed assessing laparoscopic skills using low-cost accelerometers similar to those present in smartphones, though their work was restricted to the assessment of economy of movement criteria.

Most of the existing works have focused on laparoscopic interventions that have a reduced field of motion. While a correlation between OSATS and motion analysis assessments had already been shown (Datta et al., 2001), in order to assess the complexity of the surgical skills put into practice during a procedure several criteria need to be assessed, some of which might not be covered by motion sensors only. To the best of our knowledge, there is no proposal for surgical skill assessment done with OSATS guides and motion tracking. For this reason, in this paper we present a proposal to assess surgical skills combining accelerometers and standardized assessment guide.

3. Towards an objective surgical skills assessment

In order to propose an objective assessment using an accelerometer, it was first necessary to verify the relevance of using the accelerometer to differentiate the experience level of the participants in the study, then to verify if it can be used as an assessment tool and to use it to compare the results with an assessment performed by an expert evaluator (traditional method). The proposed study is presented in the next section.

3.1. Materials and methods

General study design: This study was conducted at the School of Surgery of Nancy, *Université de Lorraine* in France. In total, 29 participants were recruited from the school's faculty of dentistry, and three groups per experience level were formed: "experienced" (n=7), "intermediate" (n=14) and "novice" (n=8). Experienced practitioners formed the first group, with a focus on oral surgery, periodontics, and implantology. The intermediate group consisted of 6th year dentistry students, while individuals without any dental surgery practice formed the novice group.

Exercise protocol: All participants receive a theoretical instructional form indicating the procedures to be performed. Each participant is asked to perform: 2 Single 0 points, 2 points in 8 and a single running suture with 4 passes of the needle on the lower jaw of a pig, the flaps of which are retracted prior to the operation (c.f. Figure 1).

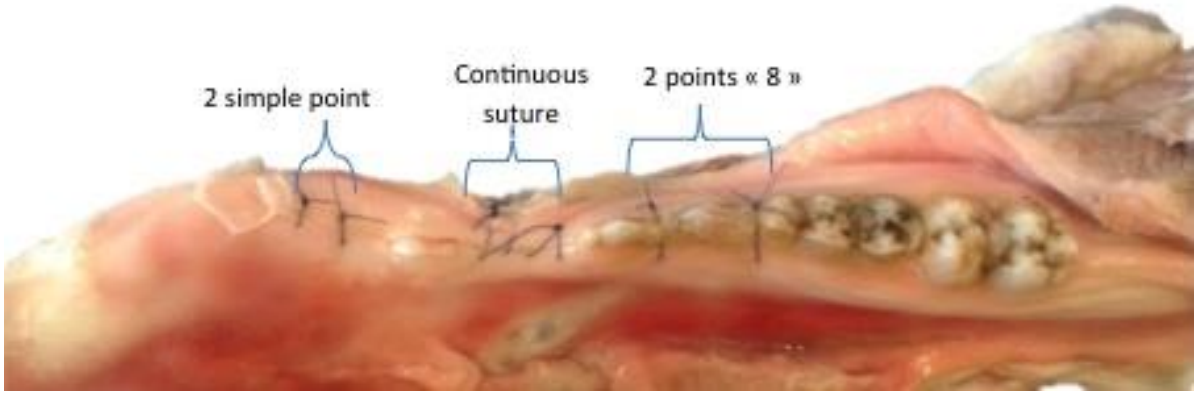


Figure 1. Protocol exercise suture points.

Motion Sensors: Two wireless motion sensors (Captiv T-sense, Tea®, Nancy), placed in the participant's wrist, were used to record the hand acceleration in the 3 axes of movement (x,y,z) at a frequency of 128 Hz. The sensors were fixed at the participants' wrists at the beginning of the exercise, the X-axis pointing to the back of the participant.

Procedure: Each participant is equipped with the motion sensors and then they are asked to perform the exercises from the protocol (suture points).

For standardized procedures, the distance between the elbow and the floor of the participants is identical for all subjects (94 cm).

Each operator is filmed anonymously throughout the completion of the procedure using a camcorder (Everio GS TD1, JVC).

Instrumentation available to achieve the sutures is identical for all participants and includes:

- Flexocrin 4.0 wire (Braun®)
- Castroviejo Clamp
- Surgical scissors
- Surgical tweezers

At the end of the procedure, a picture of the work is also taken and used to complete the evaluation of the participants.

Recording is started and stopped with the participants' hands lying flat on the table. This is a reference in terms of acceleration for the beginning and the end of the test.

Study of subjective skills: In order to have a reference evaluation of each subject's performance to compare with the information collected by the motion sensors, all the videos are broadcast on a private "youtube" link to be single blind evaluated by an experienced practitioner. Each video is evaluated by the practitioner on five criteria with a score from 1 to 5: *Respect of tissues, Time and movement, Instrument handling, Operation flow, and Knowledge of the procedure*. The score for each participant ranges between 5 and 25 points. The evaluation criteria used in the expert assessment were taken from the Hanson study of 2010 (Hanson et al., 2010), which we have completed with those of Nitsu et al. (Niitsu et al., 2013). Table 4 shows the scale used during the assessment for each criterion.

Table 4. Evaluator assessment scale.

Respect of tissue				
1 *Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments*	**2** **intermediate situation**	**3** *Careful handling of tissue but occasionally caused inadvertent damage*	**4** **intermediate situation**	**5** *Consistently handled tissue appropriately with minimal damage*
Time and motion				
1 *Many unnecessary moves*	**2** **intermediate situation**	**3** *Efficient time/motion but some unnecessary moves*	**4** **intermediate situation**	**5** *Clear economy of movement and maximum efficiency*
Instrument handling				
1 *Repeatedly makes tentative or awkward moves with instruments by inappropriate instrument*	**2** **intermediate situation**	**3** *Competent use of instruments but occasionally appeared stiff or awkward*	**4** **intermediate situation**	**5** *Fluid moves with instruments and no awkwardness*
Flow of operation				
1 *Frequently stopped operating and seemed unsure of next move*	**2** **intermediate situation**	**3** *Demonstrated some forward planning with reasonable progression of procedure*	**4** **intermediate situation**	**5** *Obviously planned course of operation with effortless flow from one move to the next*
Knowledge of specific procedure				

1	2	3	4	5
Deficient knowledge. Needed specific instruction at most steps	**intermediate situation**	*Knew all important steps of operation*	**intermediate situation**	*Demonstrated familiarity with all aspects of operation*

Objective analysis of operation flow and economy of movement: From the acceleration data provided by the motion sensors, it was possible to calculate the standard speed of the overall operation and by integrating it, a distance indicator (path length) linked to the global motion is obtained. This criterion reflects if there is an economy of movement in comparison with other candidates and can be directly compared with the "*time and movement*" criteria of the subjective evaluation.

To objectively assess the "*operation flow*" we used the total acceleration; then, using conditional formatting in R, a rating of 1 to 5 could be given to each participant for these two parameters. A detailed explanation of the calculations is given in the following section.

Statistical analysis: The results were expressed as mean ± standard deviation from the mean (m ± SEM). A one factor ANOVA or t-test two-tailed *t* was used to compare performance between groups. The frequency analyses were made using the Fischer test. Probability $P<0.05$ was considered statistically significant. For the pairs that showed a significant correlation a post-hoc HSD Tukey test was used. Analyses were made using the R language (R Foundation for Statistical Computing, Vienna, Austria) and RStudio (RStudio inc., Boston, MA, USA).

3.2. Results

3.2.1. Expert assessment of each participant video using the evaluation grid

The total scores obtained by participants range from 5-23 on a scale from 5 to 25 (cf. Table A.7). The *average scores* ranged from 14.58 ± 1.459 for the "Novice" group, 21.67 ± 0.5578 for the "Experienced" group, and 16.80 ± 0.8667 867 for the "Intermediate" group.

Regarding the *Respect of tissue* criterion, the "Intermediate" group gets an average of 3.0 ± 0.326, the "Novice" group gets 3.8 ± 0.1333, and the "Experienced" group gets the best average with 4.333 ± 0.2108. The differences are significant between the "Experienced" and "Intermediate" groups and between the "Experienced" and "Novice" groups.

In the *Economy of movement* criterion the "Novice" group gets an average of 2.917 ± 0.358, the "Intermediate" group gets a higher score with 3.20 ± 0.24, and the "Experienced" Group stands out with a better average of 4.333 ± 0.2108.

Average scores for the *Instrument handling* criterion ranges from 2.667 ± 0.284 for the group "Novice" to 4.333 ± 0.2108 for the group "Experienced", the "Intermediate" group getting an average score of 3,100 ± 0, 2769.

Flow of operation that also was objectively assessed by the sensors is denoted by an average subjective assessment of 3.083 ± 0.3128 for the "Novice" group, 3.20 ± 0.200 for the "Intermediate" group and 4.333 ± 0.2108 for the "Experienced" group.

Finally, the *Knowledge of specific* procedure criterion is estimated to average 2.917 ± 0.313 for the "Novice" group, 3.50 ± 0.3073 for the "Intermediate" group, and 4.333 ± 0.211 for the "Experienced" group.

3.2.2. Validation of the accelerometer as a differentiation tool

As previously explained, in order to validate the results it was first necessary to identify whether the accelerometer data could be used to distinguish between the expert and novice subjects. The following hypotheses were tested:

H1: Maximal accelerations in X, Y, Z axes and standard deviations are minimal and inversely related to the subject's experience.

H2: The cumulative acceleration during the procedure is inversely related to subject experience.

H3: Total procedure time is inversely related to subject experience (least time denotes more experience).

Table 5 shows the parameters that were calculated from the accelerometer data to be used in the analysis. The mean, maximal and deviation parameters corresponds to statistical features and the others are domain specific features to motion analysis. The path length feature in particular has been shown to be an important feature in surgical skill assessment (see Table 1 and Table 2). It is defined as $Path\ Length = \sum_{i=0}^{n-1} |p_{i+1} - p_i|$; since acceleration is the variation of speed over time, it is necessary to proceed to a double integration of the value in order to obtain the position eq.1 $P_t = \int \int a\ (dt)\ dt$. In this work we used the trapezoidal rule for the numerical integration with the data from the sensor.

Table 5. Movement analysis parameter definition.

Parameter	Description	Formula
Acceleration		
Resulting acceleration ($\boldsymbol{a}$)	Calculated using the Pythagoras theorem, where *x*, *y*, *z* represent the acceleration on each axis.	$a = \sqrt{x^2 + y^2 + z^2}$
Total resulting acceleration ($\boldsymbol{AT}$)	The addition of the resulting accelerations.	$AT = \sum_{i=0}^{n} a_i$
Maximal acceleration *x-axis, y-axis, z-axis*	The maximal acceleration in each axis i ($i = x, y, z$)	$a_{\max i} = max\{a_i\}$
Mean resulting acceleration ($\bar{\boldsymbol{a}}$)	The mean of the resulting acceleration	$\bar{a} = \frac{1}{n}\sum_{i=0}^{n} a_i$
Mean acceleration *x-axis, y-axis, z-axis*	The mean acceleration on each axis i ($i = x, y, z$)	$\bar{a}_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij}$
Standard deviation acceleration *x-axis, y-axis, z-axis* ($\boldsymbol{\sigma_{a,i}}$)	The standard deviation for the acceleration on each axis i ($i = x, y, z$)	$\sigma_{a,i} = \sqrt{\frac{\sum_{j=1}^{n}(a_i - \bar{a}_i)^2}{N}}$
Path Length	The cumulative hand distance based on tool movement during the intervention.	$Path\ Length = \sum_{i=0}^{n-1} \lvert p_{i+1} - p_i \rvert$
Time	Calculated using the motion sensor from the moment the hands leave the initial position until the procedure is completed and the hands return to the initial position (flat hands on the table). Sensor frequency is 128 Hz.	

Table 6 summarizes the data analysis results. Results are presented as Mean (M) ± Standard Deviation (SD). For the maximal acceleration, mean acceleration, SD acceleration no significant difference was found among the groups except for the SD acceleration x-axis for the dominant hand. However, there was a more significant difference between the groups for the Total resulting acceleration, the Total time and Total path length for both hands.

Table 6. Data analysis results.

	Expert	Intermediate	Novice	*P*
	M ± SD	**M ± SD**	**M ± SD**	
Total Time	545.524 ± 100.263	610.395 ± 148.388	841.609 ± 257.819	0.006 *
Dominant hand				
Max. Acceleration x-axis	0.922 ± 0.320	0.838 ± 0.268	0.962 ± 0.121	0.516
Max. Acceleration y-axis	0.665 ± 0.269	1.210 ± 0.707	1.080 ± 0.540	0.152
Max. Acceleration z-axis	1.187 ± 0.103	1.038 ± 0.239	1.096 ± 0.274	0.379
SD Acceleration x-axis	0.212 ± 0.076	0.229 ± 0.047	0.190 ± 0.039	0.277
SD Acceleration y-axis	0.270 ± 0.057	0.294 ± 0.065	0.311 ± 0.062	0.459
SD Acceleration z-axis	0.350 ± 0.041	0.376 ± 0.046	0.354 ± 0.078	0.504
Mean Acceleration	9.580 ± 0.028	9.627 ± 0.093	9.610 ± 0.037	0.370
Total Path Length	45.7 ± 23	85.7 ± 64.1	141 ± 110.3	0.058**
Total resulting Acceleration	668940.8 ± 122743.6	751934.7 ± 181201.9	1035054.1 ± 316235.4	0.005*
Non-dominant hand				
Max. Acceleration x-axis	0.784 ± 0.129	4.963 ± 15.623	0.883 ± 0.214	0.609
Max. Acceleration y-axis	1.944 ± 0.289	1.565 ± 0.512	2.065 ± 0.618	0.072
Max. Acceleration z-axis	1.185 ± 0.214	5.563 ± 16.775	1.238 ± 0.401	0.620
SD Acceleration x-axis	0.168 ± 0.038	0.236 ± 0.072	0.191 ± 0.042	0.041
SD Acceleration y-axis	0.241 ± 0.052	0.274 ± 0.086	0.294 ± 0.064	0.384
SD Acceleration z-axis	0.402 ± 0.039	0.403 ± 0.087	0.381 ± 0.073	0.787
Mean Acceleration	9.885 ± 0.079	9.888 ± 0.102	9.920 ± 0.095	0.708
Total Path Length	45.2 ± 16.8	72 ± 65.9	138.1 ± 148.9	0.13***
Total resulting Acceleration	689721.1 ± 123852.3	771648.4 ± 184466.6	1070392.7 ± 334148.7	0.005*

Data are presented as Mean ± Standard deviation (M ± SD)
* P <0.001, ** P <0.01, *** P <0.1 Statistically significant

In order to verify the relationship between the variables that presented a significant difference (Total time, Total resulting acceleration and Total path length) an analysis using an HSD Tukey test was performed.

The Total time factor presented a significant effect with the experience level ($p= 0.006099$). The "experienced" group ($M = 545.52 \pm 100.2$ *s*) performs the exercise in less time than the "intermediate" ($M= 610.39 \pm 148.38$ *s*) and "novice" group ($M = 841.60 \pm 257.81$ *s*). The analysis shows that the more experienced the candidate, the less time is required to perform the intervention in validation of the H3 hypothesis. Tukey's test showed that there is a significant difference between the "experienced" and the "novice" groups. Figure 2 presents the repartition of the data between total time and experience level.

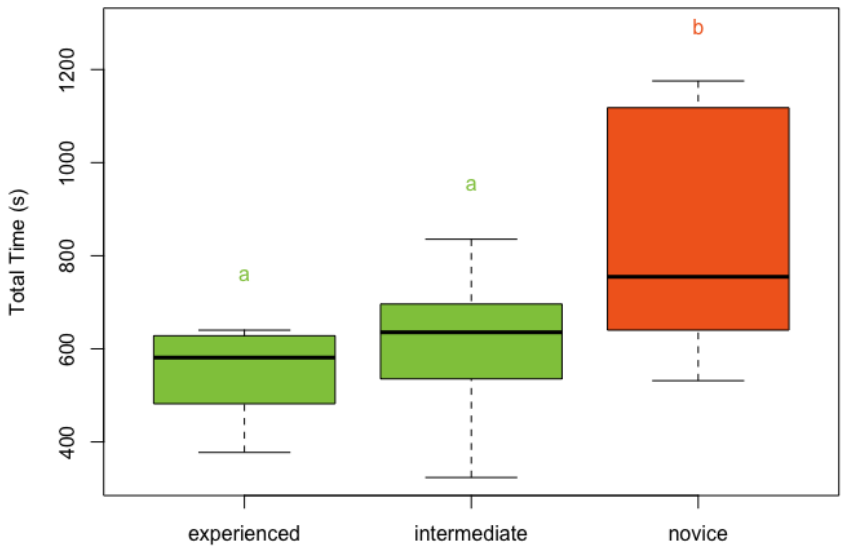


Figure 2. Total time per experience level.

In the case of the Total resulting acceleration, the statistical analysis showed that there was a significant relationship between the experience levels in each hand (*p=0.005*) for the dominant hand and (*p=0.005*) for the non-dominant hand. There is almost no difference between the hands; the "experienced" group presented a lower total acceleration than the other groups. This validates the H2 hypothesis that the least experienced participant has the highest total acceleration. Figure 3 shows the relationship between the total resulting acceleration and the experience levels. The Tukey test showed that there is an important difference between experienced and novice groups.

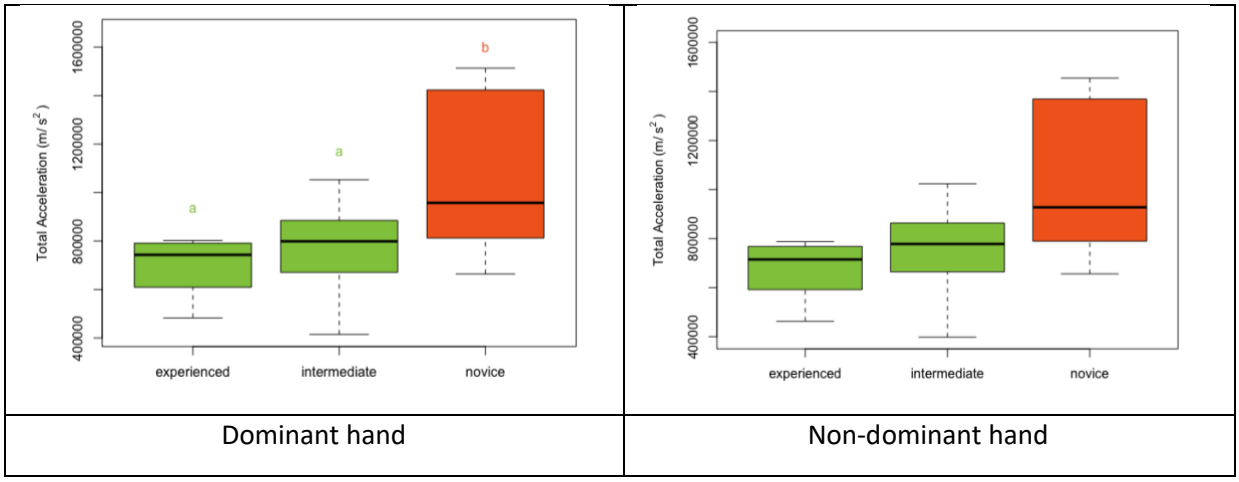


Figure 3. Total acceleration per experience group for both hands.

Finally the Total path length showed a significant difference (*p=0.058*) between experience levels for the dominant hand sensors. The "experienced" group presented a small Total length with the hands than the "novice" group this implies as in the literature that experienced practitioners had a better economy of movement. A Tukey analysis was performed; Figure 4 presents the groups' values, showing that there is an important difference that allows us to distinguish between "novice" and "experienced" groups.

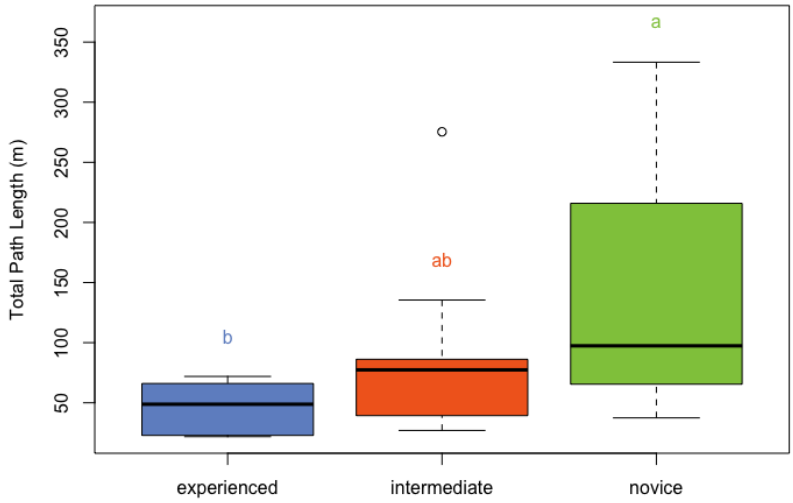


Figure 4. Total length distribution per experience level for dominant hand.

The analysis allowed us to validate the H2 and H3 hypotheses, but there was no significant difference between the maximal accelerations for each group, so there was insufficient evidence to validate the H1 hypothesis.

3.2.3. Test of the accelerometer as an objective technical skills assessment tool

Once the relationship between the motion sensor measurements and the levels of expertise had been validated, we proceeded to use those measurements as part of a skill assessment. The traditional assessment is based on the evaluation grid shown in Table 4. According to the literature, two of these five criteria can be assessed using motion analysis features (Datta et al., 2001; Godfrey, Conway, Meagher, & ÓLaighin, 2008; Sánchez et al., 2014).

The *Time and motion* criterion has been shown to correlate with the experience of the subject (cf. Table 2 metrics). According to its definition, it corresponds to the *total path length* of hand movement during the procedure.

The second criterion corresponds to the *flow of operation*; it has been shown that the number of movements and *total acceleration* are significant differentiation factors between groups (Saleh, 2008).

3.2.4. Objective grade generation

In order to compare the two assessment methods, the results obtained by the motion sensor were expressed in terms of an assessment grade (1-5). For this a five-level interval was created for each evaluated criterion (Total resulting acceleration and Path Length); the maximum and minimum values measured are used to create the intervals and then each participant is assigned a grade for each criterion. Finally, each participant's overall assessment grade (cf. Table A.7) was obtained by replacing the two criteria motion sensor grades with the evaluator grades, that is to say that the new grade in the mixed method is composed of the three grades by the evaluator (*Respect of tissue*, *Instrument handling*, and *Knowledge of specific procedure*) and the two grades generated with the motion sensors (*flow of operation* and *time and motion*). The grades corresponding to the Flow of operation criteria range from 4.43±0.53 for the "experienced" group, 3.71±0.72 for the "intermediate" group, and 2.50±1.30 for the "novice" group. For the Time and motion criteria the grades range from 5±0 for the "experienced" group, 4.50±1 for the "intermediate" group, and 3.75±1.75 for the "novice" group.

4. Discussion

4.1. Comparison between the evaluator only assessment and the mixed assessment (motion sensor included)

Results of both assessment methods were compared (cf. Table A.7). Figure 5 presents the results for each participant's grade obtained by each of the two methods (only the grades with the dominant hand are illustrated). The red dots represent the evaluator's assessment notes, and the blue dots the proposed mixed method using the motion sensors.

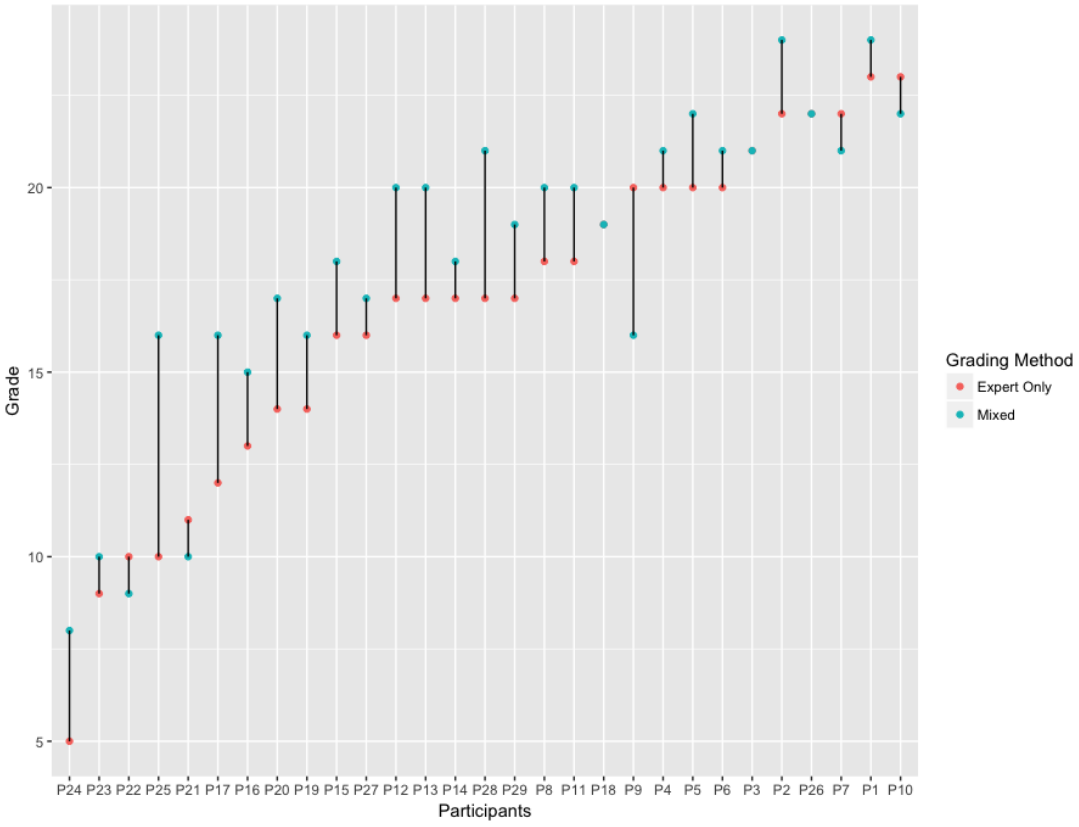


Figure 5. Evaluator and mixed assessment grades per participant.

We can observe that the grades follow the same trend, with only a few cases having a big dispersion. In comparison with the expert only evaluation, the mixed method has a precision of 72.4%. It is interesting to note that there are some participants who were initially declared in one category but in the end performed differently; in this case the grades from the mixed method and the evaluator were consistent. For example, a participant was declared as experienced but performed very poorly and was judged by the expert and the mixed method as a novice. We established that this person said it felt uncomfortable with the use of the tool in the experiment because he commonly used another type of tool.

In order to analyze the correlation among the assessment criteria a principal component analysis, PCA, was applied. Figure 6 presents a biplot chart; we can observe that all the criteria and, more importantly, the Evaluator only grade (Total) and the Mixed grade (T.Note.Capteurs) follow the same direction. Expert and intermediate participants are mostly located in the left quadrants, with the exception of the expert who performed badly. The novice categories are dispersed in the right quadrant.

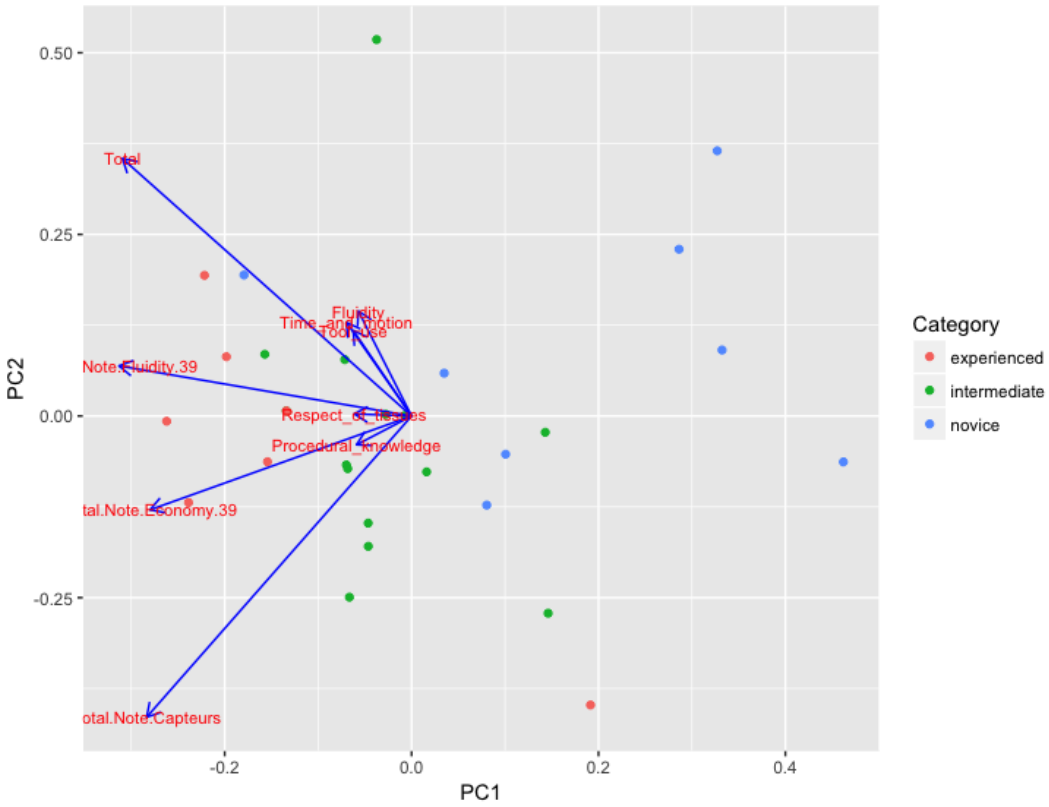


Figure 6. Principal Component Analysis biplot chart.

4.2. General discussion

Mastery of gestures and knowledge of procedures are essential to ensure maximum efficiency and safety for the patient in oral surgery and other surgical disciplines. The evaluation of skills for surgical procedures in dentistry is frequently based on a subjective assessment using qualitative criteria. To overcome this problem, various means can be implemented, such as using motion sensors. In laparoscopic surgery, sensors were successfully used to differentiate between several groups of practitioners (Salamo et al., 2014; Sánchez et al., 2014), and development of simulators with haptic technology also allows us to consider an objective assessment of future surgeons' skills (Botden, Torab, Buzink, & Jakimowicz, 2008; Dehabadi, Fernando, & Berlingieri, 2014; Joseph et al., 2014; Uemura et al., 2015). Other methods, such as tracking eyes, were also able to differentiate practitioners of different levels (Richstone et al., 2010). Finally, the use of grid corrections such as those used in the OSCE (objective structured clinical evaluation) reduces subjectivity without deleting it completely (Hanson et al., 2010; Niitsu et al., 2013; Scott et al., 2001).

The objective assessment of skills is essential in seeking certification of gesture with the ultimate goal of increasing patient safety (Grigoresco, Mignon, & Tesniere, 2014). Like laparoscopic surgery and robotic surgery, skills in oral surgery must be considered from the angle of economy of movement, which reflects a certain skill of the operator. In this study we were able to differentiate between experts and novices through the subjective evaluation of the procedure videos.

The different criteria used allowed us to classify novices and experienced practitioners. In accordance with what has been shown in robotic surgery (Perrenot et al., 2012), the group of experienced practitioners showed significantly greater economy of movement than the novice group. It therefore seems interesting to find a way to measure this parameter objectively. This is what was done in this study with the use of motion sensors. The results obtained, thanks to the use of sensors, allowed us to see a marked improvement in the experienced group compared to other groups.

We also wanted to objectively define the fluidity of the operation that characterizes the ability of the practitioner. To do this we calculated the total resulting acceleration during the procedure measured by the sensors. It was shown that the mixed grades provided by the sensors are as good as the grades provided by the evaluator.

In this study we used sensors attached to participants' wrists to avoid hindering their performance during the test. Participants did not report any difficulties with the placement of the sensors.

This study is not without limitations. Indeed, the sample size is limited to generalize the results. The model parameters need to be modified and adapted if the procedure changes. The calculation of the discrete levels is dependent on the size and performance of the groups in order to translate the measurements recorded by the motion sensors into their equivalent grades.

While an effort to keep the conditions of the test as standard as possible was made, the use of wet models implies a deterioration of the model with each procedure performed. The use of a standard model (synthetic model for example) validated for each participant could help reduce the variance that might be introduced by the manipulation of the animal model.

5. Conclusion

The use of recording sensors to objectively assess performance parameters should eventually make it possible to certify the skills of future practitioners.

In dentistry, as in other medical fields, it is necessary to identify the relevant criteria and then find a simple way to measure these criteria. Here we have studied the value and usability of a motion sensor system, allowing us to assess the economy of movement and flow of operation criteria. However, other criteria and sensors should be considered in order to achieve a completely objective evaluation. The establishment of such a system could enable self-assessment that would be relevant for students.

By combining these results with an image analysis, such as that proposed by Frischknecht (Frischknecht et al., 2013), a complete objective assessment of basic skills in oral surgery could be considered.

6. Acknowledgment

The authors would like to thank all the students and surgeons that participated in the study. In particular, the authors are grateful to the engineering students that helped run the experiments: Jessy Vincente, Maud Turin, Noel Vignacq, Adrian Carpinteiro, and Paola Romo. We would also like to thanks Fabio Cruz for its help debugging our data analysis scripts.

7. Appendix A. Participants' assessment notes for expert only and mixed methods

Table A.7. Participants' assessment with expert only and mixed method.

CODE	Category	Respect of tissues	Time and motion	Tool use	Fluidity	Procedural knowledge	Total grade (expert only)	Fluidity grade (proposed)	Economy grade (proposed)	Total grade (mixed)
1	experienced	5	5	4	4	5	**23**	5	5	**24**
10	experienced	4	5	5	5	4	**23**	4	5	**22**
2	experienced	5	4	4	4	5	**22**	5	5	**24**
26	experienced	4	4	5	5	4	**22**	4	5	**22**
7	novice	4	5	4	5	4	**22**	4	5	**21**
3	intermediate	4	4	4	4	5	**21**	4	4	**21**

5	experienced	4	4	4	4	4	**20**	5	5	**22**
6	experienced	4	4	4	4	4	**20**	4	5	**21**
9	intermediate	4	4	4	4	4	**20**	3	1	**16**
4	intermediate	4	4	4	4	4	**20**	4	5	**21**
18	intermediate	4	4	3	4	4	**19**	3	5	**19**
8	intermediate	4	3	4	4	3	**18**	4	5	**20**
11	intermediate	4	4	3	3	4	**18**	4	5	**20**
12	intermediate	4	3	4	3	3	**17**	4	5	**20**
13	intermediate	4	3	3	3	4	**17**	4	5	**20**
14	intermediate	4	3	3	3	4	**17**	3	4	**18**
28	intermediate	4	3	3	3	4	**17**	5	5	**21**
29	intermediate	4	4	3	4	2	**17**	5	5	**19**
15	intermediate	4	3	3	3	3	**16**	3	5	**18**
27	novice	3	4	3	3	3	**16**	3	5	**17**
20	novice	3	3	2	3	3	**14**	4	5	**17**
19	novice	3	3	2	3	3	**14**	3	5	**16**
16	intermediate	3	2	3	3	2	**13**	3	4	**15**
17	intermediate	3	2	1	2	4	**12**	3	5	**16**
21	novice	3	1	2	2	3	**11**	1	1	**10**
22	novice	2	2	2	3	1	**10**	3	1	**9**
25	experienced	2	1	2	2	3	**10**	4	5	**16**
23	novice	1	2	2	2	2	**9**	1	4	**10**
24	novice	1	1	1	1	1	**5**	1	4	**8**

8. References

Acton, R. D. (2015). The Evolving Role of Simulation in Teaching Surgery in Undergraduate Medical Education. *Surgical Clinics of North America*, *95*(4), 739–750. https://doi.org/10.1016/j.suc.2015.04.001

Aggarwal, R., Grantcharov, T., Moorthy, K., Milland, T., Papasavas, P., Dosis, A., … Darzi, A. (2007). An Evaluation of the Feasibility, Validity, and Reliability of Laparoscopic Skills Assessment in the Operating Room. *Annals of Surgery*, *245*(6), 992–999. https://doi.org/10.1097/01.sla.0000262780.17950.e5

Bann, S. D., Khan, M. S., & Darzi, A. W. (2003). Measurement of Surgical Dexterity Using Motion Analysis of Simple Bench Tasks. *World Journal of Surgery*, *27*(4), 390–394. https://doi.org/10.1007/s00268-002-6769-7

Bann, S., Darzi, A., Munz, Y., Kumar, B. D., & Moorthy, K. (2004). Laparoscopic virtual reality and box trainers: is one superior to the other? *Surgical Endoscopy*, *18*(3), 485–494. https://doi.org/10.1007/s00464-003-9043-7

Botden, S. M. B. I., Torab, F., Buzink, S. N., & Jakimowicz, J. J. (2008). The importance of haptic feedback in laparoscopic suturing training and the additive value of virtual reality simulation. *Surgical Endoscopy*, *22*(5), 1214–1222. https://doi.org/10.1007/s00464-007-9589-x

Brydges, R., Classen, R., Larmer, J., Xeroulis, G., & Dubrowski, A. (2006). Computer-assisted assessment of one-handed knot tying skills performed within various contexts: a construct validity study. *The American Journal of Surgery*, *192*(1), 109–113. https://doi.org/10.1016/j.amjsurg.2005.11.014

Brydges, R., Sidhu, R., Park, J., & Dubrowski, A. (2007). Construct validity of computer-assisted assessment: quantification of movement processes during a vascular anastomosis on a live porcine model. *The American Journal of Surgery*, *193*(4), 523–529. https://doi.org/10.1016/j.amjsurg.2006.08.076

D'Angelo, A. L. D., Rutherford, D. N., Ray, R. D., Laufer, S., Kwan, C., Cohen, E. R., … Pugh, C. M. (2015).

Idle time: An underdeveloped performance metric for assessing surgical skill. *American Journal of Surgery*, *209*(4), 645–651. https://doi.org/10.1016/j.amjsurg.2014.12.013

Datta, V., Chang, A., Mackay, S., & Darzi, A. (2002). The relationship between motion analysis and surgical technical assessments. *The American Journal of Surgery*, *184*(1), 70–73. https://doi.org/10.1016/S0002-9610(02)00891-7

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 model1 1No competing interests declared. *Journal of the American College of Surgeons*, *193*(5), 479–485. https://doi.org/10.1016/S1072-7515(01)01041-9

Dehabadi, M., Fernando, B., & Berlingieri, P. (2014). The use of simulation in the acquisition of laparoscopic suturing skills. *International Journal of Surgery*, *12*(4), 258–268. https://doi.org/10.1016/j.ijsu.2014.01.022

Fitts, P., & Posner, M. (1967). *Human performance*. Belmont, CA: Brooks/Cole.

Fried, G. M., & Feldman, L. S. (2008). Objective assessment of technical performance. *World Journal of Surgery*, *32*(2), 156–160. https://doi.org/10.1007/s00268-007-9143-y

Frischknecht, A. C., Kasten, S. J., Hamstra, S. J., Perkins, N. C., Gillespie, R. B., Armstrong, T. J., & Minter, R. M. (2013). The Objective Assessment of Experts' and Novices' Suturing Skills Using An Image Analysis Program. *Academic Medicine*, *88*(2), 260–264. https://doi.org/10.1097/ACM.0b013e31827c3411

Godfrey, A., Conway, R., Meagher, D., & ÓLaighin, G. (2008). Direct measurement of human movement by accelerometry. *Medical Engineering and Physics*, *30*(10), 1364–1386. https://doi.org/10.1016/j.medengphy.2008.09.005

Grigoresco, B., Mignon, A., & Tesniere, A. (2014). La simulation en santé, de la théorie à la pratique. *La Revue de L'infirmiere*, *63*(204), 17–19.

Hanson, C., Mossey, P., & Macluskey, M. (2010). The assessment of suturing skills of dental undergraduates. *European Journal of Dental Education: Official Journal of the Association for Dental Education in Europe*, *14*(2), 113–117. https://doi.org/10.1111/j.1600-0579.2009.00600.x

Joseph, D., Jehl, J., Maureira, P., Perrenot, C., Miller, N., Bravetti, P., … Tran, N. (2014). Relative Contribution of Haptic Technology to Assessment and Training in Implantology. *BioMed Research International*, *2014*, 1–9. https://doi.org/10.1155/2014/413951

Mason, J. D., Ansell, J., Warren, N., & Torkington, J. (2013). Is motion analysis a valid tool for assessing laparoscopic skill? *Surgical Endoscopy and Other Interventional Techniques*. https://doi.org/10.1007/s00464-012-2631-7

Moorthy, K., Munz, Y., Dosis, A., Bello, F., Chang, A., & Darzi, A. (2004). Bimodal assessment of laparoscopic suturing skills: Construct and concurrent validity. *Surgical Endoscopy*, *18*(11), 1608–1612. https://doi.org/10.1007/s00464-003-9312-5

Morris, B. (2005). Robotic surgery: applications, limitations, and impact on surgical education. *MedGenMed : Medscape General Medicine*, *7*(3), 72.

Niitsu, H., Hirabayashi, N., Yoshimitsu, M., Mimura, T., Taomoto, J., Sugiyama, Y., … Takiyama, W. (2013). Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surgery Today*, *43*(3), 271–275. https://doi.org/10.1007/s00595-012-0313-7

Oropesa, I., Sánchez-González, P., Lamata, P., Chmarra, M. K., Pagador, J. B., Sánchez-Margallo, J. A., … Gómez, E. J. (2011). Methods and Tools for Objective Assessment of Psychomotor Skills in Laparoscopic Surgery. *Journal of Surgical Research*, *171*(1), e81–e95. https://doi.org/10.1016/j.jss.2011.06.034

Pellegrini, C. A. (2012). Surgical education in the United States 2010: developing intellectual, technical and human values. *Updates in Surgery*, *64*(1), 1–3. https://doi.org/10.1007/s13304-011-0113-4

Perrenot, C., Perez, M., Tran, N., Jehl, J.-P., Felblinger, J., Bresler, L., & Hubert, J. (2012). The virtual reality simulator dV-Trainer(®) is a valid assessment tool for robotic surgical skills. *Surgical Endoscopy*, *26*(9),

2587–2593. https://doi.org/10.1007/s00464-012-2237-0

Ponton-Carss, A., Kortbeek, J. B., & Ma, I. W. Y. (2016). Assessment of technical and nontechnical skills in surgical residents. *The American Journal of Surgery*, *212*(5), 1011–1019. https://doi.org/10.1016/j.amjsurg.2016.03.005

Reznick, R. K., & MacRae, H. (2006). Teaching Surgical Skills — Changes in the Wind. *New England Journal of Medicine*, *355*(25), 2664–2669. https://doi.org/10.1056/NEJMra054785

Reznick, R., Regehr, G., MacRae, H., Martin, J., & McCulloch, W. (1997). Testing technical skill via an innovative "bench station" examination. *American Journal of Surgery*, *173*(3), 226–230. https://doi.org/10.1016/S0002-9610(97)89597-9

Richstone, L., Schwartz, M. J., Seideman, C., Cadeddu, J., Marshall, S., & Kavoussi, L. R. (2010). Eye Metrics as an Objective Assessment of Surgical Skill. *Annals of Surgery*, *252*(1), 177–182. https://doi.org/10.1097/SLA.0b013e3181e464fb

Salamo, O., Ismayel, A. S., Rodríguez, O., Sánchez, R., Pena, R., & Báez, V. & M. (2014). Evaluation of surgical skills for laparoscopic surgery: study of movement patterns by means of accelerometers. *VITAE Academia Biomédica Digital*, *0*(58), 4945.

Saleh, G. M. (2008). Motion Analysis as a Tool for the Evaluation of Oculoplastic Surgical Skill. *Archives of Ophthalmology*, *126*(2), 213. https://doi.org/10.1001/archophthalmol.2007.62

Sánchez, A., Rodríguez, O., Sánchez, R., Benítez, G., Pena, R., Salamo, O., & Baez, V. (2014). Laparoscopic Surgery Skills Evaluation: Analysis Based on Accelerometers. *JSLS : Journal of the Society of Laparoendoscopic Surgeons*, *18*(4), e2014.00234. https://doi.org/10.4293/JSLS.2014.00234

Scott, B. J., Evans, D. J., Drummond, J. R., Mossey, P. A., & Stirrups, D. R. (2001). An investigation into the use of a structured clinical operative test for the assessment of a clinical skill. *European Journal of Dental Education : Official Journal of the Association for Dental Education in Europe*, *5*(1), 31–7.

Shaharan, S. (2014). Evaluation of surgical training in the era of simulation. *World Journal of Gastrointestinal Endoscopy*, *6*(9), 436. https://doi.org/10.4253/wjge.v6.i9.436

Smith, S. G. T., Torkington, J., Brown, T. J., Taffinder, N. J., & Darzi, A. (2002). Motion analysis. *Surgical Endoscopy*, *16*(4), 640–645. https://doi.org/10.1007/s004640080081

Uemura, M., Yamashita, M., Tomikawa, M., Obata, S., Souzaki, R., Ieiri, S., … Hashizume, M. (2015). Objective assessment of the suture ligature method for the laparoscopic intestinal anastomosis model using a new computerized system. *Surgical Endoscopy*, *29*(2), 444–452. https://doi.org/10.1007/s00464-014-3681-9

van Hove, P. D., Tuijthof, G. J. M., Verdaasdonk, E. G. G., Stassen, L. P. S., & Dankelman, J. (2010). Objective assessment of technical surgical skills. *British Journal of Surgery*, *97*(7), 972–987. https://doi.org/10.1002/bjs.7115