RESEARCH REVIEW

Methods and Tools for Objective Assessment of Psychomotor Skills in Laparoscopic Surgery

Ignacio Oropesa, B.Sc.,*,†,[1] Patricia Sánchez-González, M.Sc.,*,† Pablo Lamata, Ph.D.,* Magdalena K. Chmarra, Ph.D.,‡ José B. Pagador, B.Sc.,§ Juan A. Sánchez-Margallo, B.Sc.,§ Francisco M. Sánchez-Margallo, Ph.D.,§ and Enrique J. Gómez, Ph.D.*,†

**Bioengineering and Telemedicine Centre (GBT), ETSI Telecomunicación, Universidad Politécnica de Madrid (UPM), Madrid, Spain; †Networking Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Zaragoza, Spain; ‡Department of BioMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering (3mE), Delft University of Technology, Delft, The Netherlands; and §Minimally Invasive Surgery Centre Jesús Usón, Cáceres, Spain*

Training and assessment paradigms for laparoscopic surgical skills are evolving from traditional mentor–trainee tutorship towards structured, more objective and safer programs. Accreditation of surgeons requires reaching a consensus on metrics and tasks used to assess surgeons' psychomotor skills. Ongoing development of tracking systems and software solutions has allowed for the expansion of novel training and assessment means in laparoscopy. The current challenge is to adapt and include these systems within training programs, and to exploit their possibilities for evaluation purposes. This paper describes the state of the art in research on measuring and assessing psychomotor laparoscopic skills. It gives an overview on tracking systems as well as on metrics and advanced statistical and machine learning techniques employed for evaluation purposes. The later ones have a potential to be used as an aid in deciding on the surgical competence level, which is an important aspect when accreditation of the surgeons in particular, and patient safety in general, are considered. The prospective of these methods and tools make them complementary means for surgical assessment of motor skills, especially in the early stages of training. Successful examples such as the Fundamentals of Laparoscopic Surgery should help drive a paradigm change to structured curricula based on objective parameters. These may improve the accreditation of new surgeons, as well as optimize their already overloaded training schedules.

Key Words: **laparoscopic surgery; objective evaluation; accreditation; metrics; basic skills; virtual reality; human motion tracking; classification.**

INTRODUCTION

Minimally invasive surgery (MIS) has changed the way surgery is performed in operating rooms (OR). In many procedures, it has become the recommended standard, displacing open surgery [1]. Laparoscopy, one of the most common MIS approaches, has been adopted by several surgical sub-specialties, including gastrointestinal, gynecologic, and urologic surgeries [2]. Effective training and assessment of surgeons in these new techniques have become one of the major concerns of hospitals and clinics in recent years, fuelled mostly by patients' and society's demands for safer surgeries and well prepared physicians [3–6].

Traditional Halsted-based training [7] is potentially unsafe for the patient and, as a consequence, no longer ethically sustainable [4]. There is a tendency to move the early phases of training, concerned with the acquisition of motor skills, outside the OR. For this reason, laboratory settings including, for example, box trainers and/or virtual reality (VR) simulators have been developed [4].

Another issue that has become evident is the need for structured formation programs inside the OR [8]. Assessment based on In-Training Evaluation Reports (ITERs) [9] is subjective, expensive, and prone to

[1] To whom correspondence and reprint requests should be addressed at Bioengineering and Telemedicine Centre (GBT), ETSI Telecomunicación, Universidad Politécnica de Madrid (UPM). Avda Complutense, 30, 28040, Madrid, Spain. E-mail: ioropesa@gbt.tfo.upm.es.

undesirable side effects, such as a halo effect [10]. The halo effect is a result of influenced perception of the performance in one area (e.g., laparoscopic task) by the performance in another area (e.g., relationship between trainee and mentor) [11]. Additionally the ITERs are periodically written, depriving the trainee of immediate feedback, and are subject to the evaluators' long-term memory [12].

Structured reports, based on checklists and immediate end-product analysis have been proposed and validated [11,13–16]. The Objective Structured Clinical Examination (OSCE) [13] has been used to assess clinical performance of trainees on various clinical stations (comprising different tasks). Since OSCE mainly focuses on the assessment of procedural knowledge and attitude of the trainee towards the patient, technical evaluation of psychomotor skills is not given leading relevance [9].

The Objective Structured Assessment of Technical Skills (OSATS) pays more attention to motor skills assessment [14]. Validity has been fully established for skills assessment ranging from simple tasks to advanced, complex procedures [15]. Implementation of the OSATS in the OR, however, may present ambiguities in the scoring systems [16]. Another drawback is the high amount of resources required, from the number of experts deployed at each station to evaluate the trainees, to the marginal costs of each exam per candidate [9]. Laparoscopic video offline-evaluation has been proposed to reduce some of these costs with good reliability results [17]. However, the presence of a reviewer is still required, and trainees do not obtain immediate feedback about their performance. A counterpart of OSATS for MIS—the Global Operative Assessment of Laparoscopic Skills (GOALS)—has been developed by Vassiliou *et al.* [11]. GOALS are not procedure-specific reports, and as such, they can be used for any MIS procedure.

As the need for objective and structured assessment of technical performance grows, new tools and methods have emerged over the last few years [9, 15, 18]. Training methods are being gradually changed, leaving the traditional ways behind on behalf of criterion-based curricula [19]. This tendency has been favored by the development and advances on tracking systems and computing technologies, which have led to the appearance of human motion tracking devices and virtual simulation for surgical training [20]. With new ways to measure surgical performance, the remaining question is: "What does it mean to be a competent surgeon?" Thus, simultaneous to the technological advances, much research has been devoted to the development of metrics for skills assessment. These metrics determine to a great extent a measuring instrument's proficiency, and are necessary to provide evidence of its reliability and validity as an assessment tool [8] (Table 1).

The purpose of this review is to present a state-of-the-art on the new tools available for acquisition and analysis of information concerning surgical performance, and their influence in the development of new accreditation programs. For this end, we have modeled the process of surgical assessment as a three-sided problem: (1) the clinical side; (2) the technological side; and (3) the analytical side [21]. The clinical side deals with the definition of the optimal tasks, metrics and conditions to consider for the assessment of the different psychomotor skills required. The technological side is related to the use of tracking technologies and/or computer assisted systems for the creation of surgical training and assessment environments that allow capturing objective data concerning a surgeon's skill. Finally, the analytical side studies the use of statistical analysis and machine learning algorithms for data analysis, in order to ascertain whether automatic classification systems to aide surgical assessment are viable or not.

METHODOLOGY

Search of the literature was performed using PubMed and Google Scholar public databases. Key words employed were: "laparoscopy", "minimally invasive surgery," "surgical assessment", "psychomotor skills", "objective evaluation", "validation", "virtual reality, " and "motion analysis." Obtained article' bibliographies were also checked for new references. Additionally, validation brochures for commercial virtual reality simulators were scanned. For all considered sections, articles not related with laparoscopic surgery were filtered. No *a priori* restrictions regarding publishing date or language were applied.

For the Clinical Definition of Metrics section, recovered results were scrutinized and filtered by construct validation studies employing objective data. Our purpose was to cross-reference the most recurrent metrics and surgical tasks for box trainers or VR simulators, identifying which parameters influence on the different psychomotor skills. In this way, it was possible to discern general patterns where a given metric/set of metrics yields significant differences between training groups, with respect to different tasks and abilities. Thus, only reports showing positive results are included in this review. A complete up-to-date detailed revision on valid/nonvalid systems is presented in [18].

In the Tracking Technologies for Skills Assessment section, a technological overview of tracking tools for skill assessment is given, from an application point of view. Depending on the setting on which the tracking systems are used, we propose to categorize them as those used in (1) VR simulated environments, and (2) real settings. The first one refers to tracking

TABLE 1

Validation Requirements for an Assessment Tool

Validation parameters	
Face	Subjective expert review of the test contents
Content	Subjective detailed examination of the test contents
Construct	Objective measurement. Degree to which the test captures the hypothetical quality it was designed to measure
Concurrent	Objective measurement. Extent to which scores on a test and a control instrument are correlated
Predictive	Objective measurement. Extent to which scores predict actual true performance
Reliability parameters	
Inter-rater	Extent to which two different evaluators give the same score in a test performed by a user
Intra-rater	Internal consistency of an evaluator when grading on a given test on different occasions
Test-retest	Extent to which two different tests made by the same person in two different time frames give the same result

technologies applied on software-generated environments ("mimicked" training)—VR simulators. The latter one is applied on real settings, mainly box trainers, but also on mannequins, cadavers, animals, and during real OR interventions. Throughout the text, these will be referred to as human motion tracking (HMT) systems. More detailed description on the technical aspects of surgical tracking devices can be found in [20].

In the Automatic Analysis for Skills Assessment section, a review is given of the studies in the field of automatic classification of surgical skills, highlighting advantages but also the need for further clinical evidence regarding their effectiveness. Articles were screened for those reporting the use of high-level statistical analysis and machine learning techniques for laparoscopic surgical assessment of competence based on motion and force data.

CLINICAL DEFINITION OF METRICS

Research has been devoted to the definition and validation of new metrics for performance assessment [22, 23], as well as to the determination of the ideal tasks and skills to train [8]. Task definition is usually setting-dependant and, as such, can vary whether if built for a box trainer or for a virtual simulator. We have broadly categorized surgical tasks in the classes shown in Table 2, following the laparoscopic skill taxonomy presented by Lamata [23].

A total of 32 studies cross-referencing tasks and valid metrics, both in box trainers and virtual reality simulators, were selected [24–56]. Metrics have been classified according to the taxonomy proposed by Fried and Feldman into two main categories: (1) efficiency metrics and (2) quality metrics [12]. A brief description of the considered metrics is shown in Table 3 [16]. Results from the cross-validation study are presented in Table 4.

Efficiency metrics are related with measurable physical parameters [12]. Their definition is usually precise and supported by a strong theoretical background. These metrics require the use of tracking devices in order to be acquired. In consequence, they are objective, reproducible and little prone to misinterpretation. A distinction can be made between (a) motion-derived and (b) force-derived metrics.

For all considered tasks, the most prominent motion efficiency metrics are time, path length and economy of movements [24, 30, 31, 34]. They are especially important in the tasks which involve touching or grasping objects on the scenario, both with one or two instruments; and that require bi-manual coordination. A stipulated reasoning behind this is that an expert surgeon performs a task more swiftly, denoting a more clear perception of the surgical space and the required strategic approach [51]. However, depending on the nature and difficultness of the task, these metrics might not always show significant differences [57]. For these same tasks, a number of studies have shown that speed of movements can be a differentiating aspect [29, 31, 48]. This may partly be due to the difficulty of determining an optimal path, which in any case will depend on the task's goal. It is a common interpretation to consider that the straight line between two points is the ideal path. In laparoscopic surgery, however, it has been proved that this is not always the case [58]. Motion smoothness is also reported as a determining factor for manipulation tasks, such as those involving object grasping, transfer, cutting, or suturing [25, 26, 46]. Other motion metrics relate to the trainee's mastery of the task space, in terms of dimensions and orientation. Those include: depth, angular area, volume, and spatial perception. However, it is difficult to establish their validity since they are not often considered. Nevertheless, the studies which have featured them have shown some significant differences for tasks such as grasping, bi-manual coordination, clipping, cutting and navigation [37–40, 44, 46].

Force-related metrics have been mostly validated for suturing tasks [49, 50, 56]. Additionally, Rosen *et al.* demonstrate that experienced surgeons apply higher

TABLE 2

Surgical Metrics Featured in the Literature

Efficiency metrics	
Time	Total time to perform a task (s)
Path length	Total path followed by the laparoscopic instrument (m)
Economy of movements (EOM)	Shortest distance to accomplish task/total distance (%)
Economy of diathermy	For diathermy: excess burn time/optimal burn time (%)
Speed	Rate of change of the instruments' position per second (m/s)
Motion smoothness	Abrupt changes in acceleration resulting in jerky movements of the instruments (m/s^3)
Instrument orientation	Amount of instrument rotation, measures the ability of correctly placing the instrument (rad)
Depth	Total path length traveled in the instrument's axis direction (m)
Angular path	Sum of all angular paths about the instrument's pivot point (°)
Angular area	Area between the farthest positions occupied by the instrument in the camera plain (rad^2)
Volume	Angular area x Depth ($rad^2 \cdot m$)
Force/torque	Instrument – tissue force (N) and torsion ($N \cdot m$) interactions

Quality metrics	
Outcome	Final score of the task performed. Task-dependant
Errors	Errors performed during the task. Task dependant
Idle states	Time periods when instrument movements/interactions are minimal
Task repetitions	Number of repetitions required on a task before achieving satisfactory completion
Collisions/tissue damage	On VR: detection of incorrect collisions and damage performed to background tissues

force/torque magnitudes during tissue dissection than novices, and vice versa for tissue manipulation during full procedures [59]. More recently, Horeman *et al.* have approached instrument/tissue forces detection by means of a pressure platform placed under the box trainer task [56].

Quality metrics relate to a task's definition and execution [12]. It has been demonstrated that neither one performance measure nor efficiency metrics alone adequately measure competence, since competence is multi-factorial in nature, with knowledge, judgment, behavior, and technical abilities each playing a major role [60, 61]. Therefore, various performance measures that are essential for surgical competence need to be taken into account. Most prominent amongst these metrics are the errors performed and the end-product analysis. These are extensively validated throughout all tasks, and denote a further understanding of a trainee's true skill level [30, 33, 36, 38, 40]. VR simulators allow quantifying tissue damage by collisions' detection, which can be seen as an indicator of the spatial perception of the environment [23]. Where present, this has generated positive validation results as a differentiating metric [24, 36, 40]. Another quality metric—detection of idle states—has not gained much attention. Nevertheless, studies show significant differences between a novice surgeon and a more experienced surgeon on suture tasks and more complex procedural chores [49, 50]. The principle behind this metric is that a novice surgeon takes more time to plan the next move than a more experienced surgeon.

TRACKING TECHNOLOGIES FOR SKILLS ASSESSMENT

The application of tracking technologies to MIS skills assessment allows the registration of efficiency metrics. These technologies usually employ optical,

TABLE 3

Classification of Laparoscopic Basic Tasks

Task classes	
Touch/coordination	Involves touching fixed or mobile targets with the tip of the instrument
Navigation	Navigating the camera within the scenario
Peg grasping	Grasping and placing a target in a predefined point of the scenario
Peg transfer	Grasping tasks involving tool transfer of an object
Navigation + touch/grasp	Navigating the camera while manipulating objects/targets
Bi-manual coordination	Bi-manual coordinated manipulation of miscellaneous objects within the scenario
Cutting/dissection	Fine tuning of cutting and dissecting skills
Clipping	Placing a clip on a target
Cauterization	Cauterizing target points on a task
Suture	Suturing and knotting skills

TABLE 4

Correlation of Basic Tasks and Validated Metrics

	Basic tasks									
Metrics	Touch/ coordination	Navigation	Navigation + touch/grasp	Peg grasping	Peg transfer	Bi-manual coordination	Dissection	Clipping	Cauterization	Suture
Efficiency										
Time	[24] [29] [30] [34] [36] [38] [39] [40] [47] [48]	[27] [30] [31] [33] [34] [39]	[24] [33] [34] [36] [38] [39] [40]	[24] [26] [33] [34] [36] [38] [40] [42] [45] [46]	[30] [31] [33] [42] [43] [46]	[24] [33] [34] [35] [36] [38] [40] [46]	[24] [28] [30] [34] [36] [38] [39] [40] [46]	[30] [31] [34] [37] [38]	[30] [31] [43]	[25] [26] [36] [44] [45] [50] [51] [52] [53] [54]
Path length	[29] [34] [38] [40]	[34] [39] [40]	[33] [34] [39] [40]	[26] [33] [34] [40] [45] [46]	[31] [33]	[34] [36] [37] [38] [40] [46]	[26] [28] [34] [36] [37] [40] [46]	[34] [36] [37]	[46]	[25] [27] [36] [44] [45] [50] [53] [54]
Economy of movements	[24] [29] [34] [47] [48]	[24] [27] [34]	[24] [34]	[34] [38] [42]	[31] [42] [43]	[34] [37] [38]	[34]	[34]	[42] [44]	[51] [52] [54]
Economy of diathermy									[42]	
Speed	[29]* [48]†				[31]*			[31]*		
Motion smoothness		[26]		[26] [46]	[46]	[46]	[26] [28] [46]			[25] [26]
Instrument orientation				[51]						[44] [50]
Depth				[45][46]		[46]	[46]			[44] [50]
Angular path	[40]	[39] [40]	[39] [40]	[38] [40]		[37] [38] [40]	[40]	[37]		
Angular area				[46]						
Volume				[46]						
Force/torque										[49] [50] [56]
Quality										
Outcome	[30] [35]	[30] [31]	[24]	[36] [45]	[30] [32]	[24] [36] [37] [55]	[27] [30] [32] [36] [37]	[30] [32] [36] [37]	[30] [32] [44]	[27] [36] [44] [46] [51] [53]
Errors	[40]	[40]	[33] [35] [39] [40]	[24] [34] [38] [40]	[33] [43]	[24] [35] [38] [40]	[26] [34] [35] [39] [40]	[34] [35] [36] [38]	[43] [44]	[36]
Idle states										[49] [50]
Task repetitions				[41]			[41]	[41]		
Collisions/tissue damage	[39] [40]	[40]	[24] [40]	[40]		[36] [37] [40]	[40]			[40]

*Average speed.

†Instantaneous speed.

electromagnetic, mechanical, or ultrasonic tracking of the instruments/hands movements, and can indistinctly be used on several training settings, whether virtual or real.

Tracking Systems in Virtual Reality Simulators

Virtual simulation has become a major trend in the field of surgical training, and many attempts have been made to develop and validate diverse commercial and research models, as shown in a recent meta-analysis performed by Gurusamy *et al.* [62]. VR simulators offer various advantages that are valuable for the training and assessment of motor skills. They allow for training in controlled environments, are generally available for the trainee, and do not require the presence of a supervisor (which may help to optimize the mentors' schedules) [63].

The different VR simulators identified for this review are shown in Table 5 [24–44], [64–77]. VR simulators can be differenced by the way they make use of their didactic resources (Fig. 1) [78]. On the one hand, simulators such as MIST-VR's "Core Skill" module (Mentice AB, Göteborg, Sweden) or SIMENDO (Simendo, Rotterdam, The Netherlands) focus on formation of psychomotor skills rather than on developing complex anatomical scenarios. On the other, VR simulators such as LapMentor (Simbionix, Lod, Israel) or LapSim (Surgical Science Ltd., Göteborg, Sweden) adopt the trend of simple task training, but make a wider use of available computer resources (e.g., realistic scenarios [23], force feedback [79, 80]) to enhance user interaction.

The use of VR simulators for psychomotor skills assessment greatly benefits from two factors:

- They allow the acquisition of both efficiency (motion, force) and quality metrics. This is achieved thanks to the combination of tracking technologies and the computer-generated environments, which make possible not only tracking the instruments, but also having control over all the elements in the scenario. Thus, objective quantification of error-counts, repetitions or end-product analysis, for example, is viable.
- They provide immediate feedback to the trainee based on these metrics, as well as keeping an updated version of his/her learning curve, which enhances the learning process [81].

Despite the advantages of VR simulators, a number of limitations have influenced their clinical implementation [23]. For example, there are resource-derived constraints, such as simulators costs, and trainees' loaded schedules, which leave them little time for practice. There are cases in which these advanced and sophisticated systems are available in the hospital but residents do not find the time or motivation to use them for training. VR realism and interaction might not be critical for their didactic value, but are often

TABLE 5

VR Simulators: Principal Models, Characteristics and Positive Validation Studies

Simulator	Module	Scenarios	Environments	Force feedback	Construct validity	Concurrent validity	Predictive validity
MIST-VR	Core Skills	Simple tasks	Non-anatomical	No	[42] [43] [44]	[42] OR	[65] [66] [67] [68]
	Nephrectomy	Procedure	Anatomical	Yes	–	–	–
LapMentor	Basic/Essential/Suture	Simple/advanced tasks	Non-anatomical	Yes	[29] [30] [32]	[69] GOALS	[70]
	Lap Chole/Ventral Hernia/ Gastric Bypass/ Gynecology/ Sigmoidectomy"	Procedures	Anatomical		[31]		
LapSim	Basic Skills	Simple/ advanced tasks	Both	Supported	[34] [35] [36] [37] [38] [40] [41]	[71] BT [72] BT	[73]
	Cholecistectomy/Gyn/ Appendectomy	Procedure	Anatomical		[39]		
	Suturing & Anastomosis	Advanced tasks	Non-anatomical		[37]		
ProMIS	Basic Skills	Simple tasks	Hybrid Simulator	Realistic Interaction	[25] [26] [27] [28]	[74] OR [75] LapSim	[76]
	Procedures	Procedural tasks			[64]		
SIMENDO	–	Simple/ advanced tasks	Non-anatomical	No	[33]	[77] BT	–
SINERGIA	–	Simple tasks	Both	Supported	[24]	–	–

Concurrent validation results show the method used for comparison.
BT = box trainer; OR = operating room.

the key elements to gain the acceptance of physicians. Moreover, there are mentality-driven constraints, such as thinking of a surgical simulator as a videogame, which has no didactic value. Prior experience with videogames can also be a handicap when facing virtual simulators [82]. It can even happen that such systems will create a false sense of security, built on the development of incorrect habits while getting used to a virtual environment.

Academic research efforts continue exploring the boundaries of the capabilities of VR technologies. As an example, the SINERGIA laparoscopic simulator explored the area of perceptual skills [83, 84] . The simulator was conceived as a means for training and assessment of motor skills on the first stages of surgical formation, comprising seven didactic units (Fig. 2).

The simulator features advanced assessment feedback, by means of an objective evaluation module that allows monitoring of trainees' learning curves [85]. In order to manage all evaluation data, different graphics modalities are implemented for easy following and understanding of the surgical skills' evolution of the trainee. Comparisons between different individuals or groups of pupils and data exchange for statistical analysis are possible in the user interface. Objective metrics' definition allows trainees to learn from their mistakes by means of indications when errors are performed (formative feedback) or by visualization of the global practice score (summative feedback).

Human Motion Tracking in Real Settings

HMT technologies provide the means to capture and register efficiency metrics of the surgeon's performance in real environments. In contrast to VR simulators, they offer a cheaper tracking alternative that can be used in almost every training setting, from box trainers to the OR. Quality measures are, however, more complicated in these systems. Measurement of quality parameters has been approached in a number of studies by using additional sensors in the training scenario [86, 87]. Typically, this assessment requires the presence of a trained supervisor, and the definition of clear, structured checklists are necessary [28, 53].

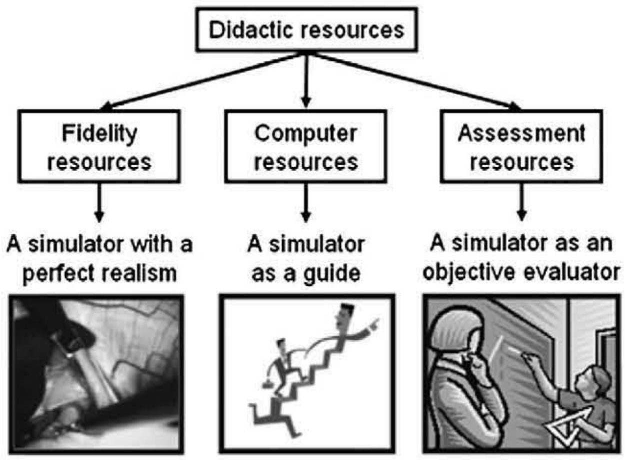

FIG. 1. The three concepts of a VR surgical simulator driven by the use of different didactic resources [78]. Author: Biomedicine and Telemedicine Centre (GBT).

Table 6 shows the most widespread HMT capture systems for surgical assessment identified in the literature [44–55, 88–91]. Their classification can be done according to the nature of the acquired information. In a majority, they are used to record both the position (x, y, z coordinates) and orientation (yaw, pitch and roll) of the surgical instruments. In this review, the following systems are considered: ADEPT [92], ICSAD [52] (which measures hand movements rather than instrument's position), CELTS [45], Zebris [48], TrEndo [93], ARH [94], HUESAD [47] and BlueDRAGON (which also measures force parameters) [95].

Most frequently, HMT systems are used on box trainers, where the performance of surgeons in basic and advanced reproducible tasks is recorded [45–48, 53]. This trend can easily be explained, since there is a need to certify a surgeon's psychomotor skills before allowing him/her to operate on a patient. However, active sensing, which relies on sensors mounted on the surgical instrument, presents the disadvantage of introducing new elements on the surgical theatre, thus altering it; and also of modifying the instruments' ergonomics, and therefore can be a drawback for OR inclusion. Also, not all HMT devices are suited for portability: the bulkiness and configuration of some of them make them fit only for a closed-up number of box-trainer tasks, and compromises their possible migration to the OR [20]. Exceptions to this are BlueDRAGON, which has been reported to be used on live animals [49], and ICSAD, validated while performing surgeries on real patients with the help of video assistance [96, 97].

An alternative approach for passive and unobtrusive tracking is the analysis of the laparoscopic videos, which allows registering movements employing computer vision techniques [98]. The non-invasive characteristics of an assessment system based on these technologies makes it fit for any potential training scenario inside and outside the OR. This way, information about position and trajectories can be acquired. The concept is not new, and has already been exploited in the ProMIS simulator (Haptica, Dublin, Ireland). Tracking 3D movements of MIS instruments in the ProMIS requires two cameras for stereoscopic vision [25]. The challenge, however, is to do the same thing using 2D information obtained using one camera only (endoscope). This idea has already been pursued on other

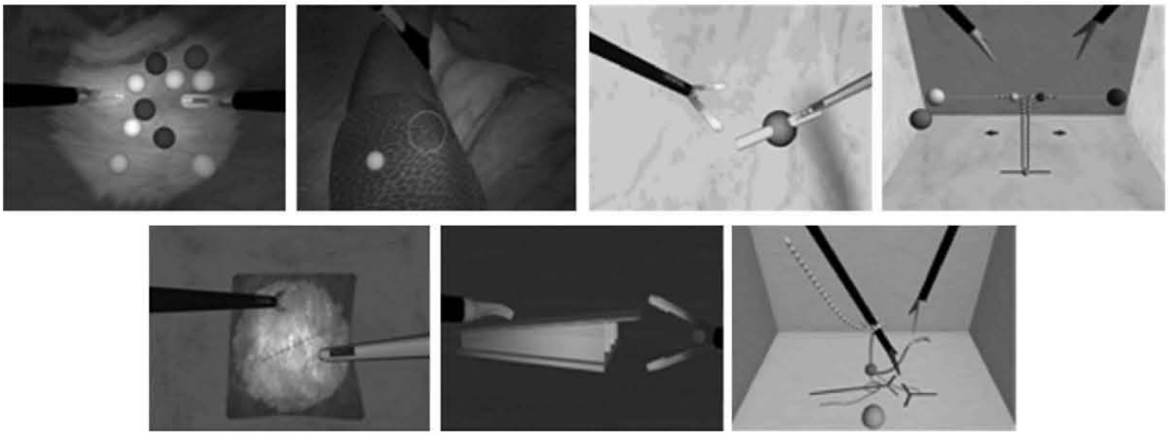

FIG. 2. An example of VR training solution and its didactic units, the SINERGIA simulator. From left to right, top row: coordination, navigation, grasping, pulling; bottom row: cutting, dissection, suturing. Author: SINERGIA Consortium.

application fields such as robotic surgery and navigation [99, 100].

Several ways to extract the position of the surgical instrument are being currently researched. 2D tracking has been validated for assessment of basic laparoscopic tasks (eye-hand coordination task, camera manipulation task and two-handed manoeuvres) on box trainers [101, 102]. 3D pose extraction has been proven feasible, based solely on the instrument's diameter and the endoscope's field of view, and validation studies as an assessment tool are currently being performed [103].

Laparoscopic video analysis for surgical skills assessment presents itself as an interesting alternative to active based tracking devices: it provides the means to calculate a wide range of motion efficiency metrics in a non-intrusive way: surgical instruments are not modified and thus the ergonomic experience of the trainee is not altered. Additionally, it offers a cheap and portable assessment tool that can be either exploited on laboratory settings such as box trainers or on real procedures in the OR.

AUTOMATIC ANALYSIS FOR SKILLS ASSESSMENT

One of the key points in a training program is providing trainees with immediate feedback of their performance [23]. VR simulators and HMT systems are able to provide real-time scores in the form of evidence-based reports (path length, errors, overall score, time, etc.) [9]. However, this information is devoid of meaning without proper interpretation which, if provided by a tutor figure, may lose its character of immediate and objective.

Constructive feedback can be provided in many ways. It can be in the form of simple reports based on weighted averages such as proposed by Cotin *et al.* [22]. It may also be handed by informative messages where, next to measurements like time or movements, formative assessment is handed with what could be the advice of a surgical expert, with messages like "too much tissue bitten" [23, 104]. Or it can be provided by a progress monitoring software, which allows trainees to view their achievements and compare their learning curve with those of other residents [85, 96, 97].

TABLE 6

HMT Devices: Principal Models, Characteristics, and Positive Validation Studies

HMT systems	Technological base	Range of application	Metrics registered	Portability	Construct validity	Concurrent validity
ICSAD	Electromagnetic	BT, OR*	Hand movements	Yes	[51] [52] [53] [54]	[89] OSATS [90] OSATS
ARH	Electromagnetic	BT	Motion	Yes	[88]	–
BlueDRAGON	Mechanical	BT, OR†	Motion/Force	No	[49] [50]	–
CELTS	Mechanical	BT, VR	Motion	Yes	[44] [45]	–
Adept	Mechanical	BT	Motion	No	[55]	[91] OR
Zebris	Ultrasound	BT	Motion	Yes	[48]	–
HUESAD	Optical	BT	Motion	–	[47]	–
TrEndo	Optical	BT, VR	Motion	Yes	[46]	–

Concurrent validation results show the method used for comparison.
BT = box trainer; OR = operating room; VR = virtual reality simulator.
*OR real patients.
†OR animals.

Establishing surgical competence is another crucial aspect for surgical assessment [16]. High level statistical analysis and machine learning techniques are employed to infer knowledge and correlate metrics information to surgical expertise, so automatic classification of trainees according to their competence is attained. These methods usually require two phases: (1) training and (2) classification [105]. During the training phase, representative data from each surgical level (e.g., novices, intermediates, experts) is used to establish the different classes representative of the competence level. In the classification phase, the data recovered for a new surgeon is compared to those classes, and an assignation to one of them is performed based on a likelihood probability.

Several techniques have shown promising results for surgical assessment (Table 7). Sequential analysis of surgical tasks by Markov modeling has been the most common approach [49, 50, 106–111]. Simple Markov models (MM) interpret sequences of actions as a series of steps, defined by a closed number of states and the probabilities from going from one state to another. States are usually taken as sets of predefined surgical maneuvers (or surgemes) [106]. MMs that decompose tasks into states based on force/torque patterns have been employed by Rosen *et al.* for modeling surgical steps with an 86% success rate [50]. However, definition of these surgemes is often done manually and, as an alternative, hidden Markov models (HMM) have been widely used [49, 106–111]. Rather than requiring a previous definition of actions, HMMs model their states as probabilistic functions based on physical observations (such as motion metrics) [112]. Thus, they are powerful tools for classification without previous surgical knowledge, whether at (1) task level [49, 107–109] or (2) surgeme level [110, 111]. Results ranging in success rates between 90-100% (Table 7) show their potential usefulness for competence assessment. Despite this, determining the number of states to use and the model's topology can be a daunting task, subjected in many cases to trial and error [107, 112]. Moreover, reported results seem not to be so different between using MMs or HMMs [106], although HMMs flexibility may help to better accommodate the needs of different assessment systems.

Nonsequential analysis has also been approached, generally basing competence level on a combination of metrics. When considering several parameters, however, it is sometimes desirable to reduce the dimensionality of the problem. Several authors have employed methodologies based on linear discriminant analysis (LDA). LDA correlates information from different metrics to detect redundancies, and additionally can perform classification based on the resulting simplified data. Chmarra *et al.* explored the use of LDA combined with principal component analysis (PCA) as a way to compare performance on a box trainer for novices, intermediates, and experts using six different motion metrics [46]. Results yielded a success rate of 74% on a first report. Lin *et al.* combined the use of LDA with a Bayes classifier for a suture task, obtaining success rates of 92% [113]. Finally, fuzzy logic, which allows establishing patterns of knowledge based on the training data, has been proposed in [114, 115]. However, reported results did not fully meet the expectations for determining surgical competence.

Choosing a classification scheme for automatic surgical assessment is not a trivial job. Sequential analysis based on Markov modeling has been proved to be robust and accurate. However, its complexity and computational requirements may handicap its use. Moreover, the fact that previous expert surgical knowledge must be considered is an additional potential source of error. In this sense HMMs are key techniques, proven both by their recurrence and results. Their usefulness for surgery modeling has been exploited successfully in other surgical fields such as robotics [116, 117]. The differences with simple MM in terms of results are not so significant, though their nature makes HMMs more flexible to the requirements of competence assessment. Finally, LDA also shows potential as a less complex yet powerful classifier. The reports, however, are still scarce, and additional validation is necessary. In general, studies presented here are limited to certain tasks and metrics, and limited in the number of participants. Therefore, further proof of validation must be given. Also, exploration of other techniques such as neural networks or support vector machines should, furthermore, be explored.

DISCUSSION

New tools and techniques developed for the benefit of surgical training have been presented. Based on our experience, we believe that they have the potential to be adapted into surgical training programs, fulfilling a complimentary role on the evaluation of motor skills.

There is a crescent pressure to develop training programs that on one hand address social pressure to reduce clinical errors, and on the other adjust to the time constraints affecting both trainees and mentors [2, 118]. The general consensus behind these programs is that the moment when the trainee confronts a real surgical procedure should be postponed as much as possible, leaving the acquisition of the basic cognitive and psychomotor skills on stress-free environments, where the learning process can be more effective without compromising patient safety. Thus, according to [2], an integral formation program should be structured in four major levels (Fig. 3): (1) training of basic and advanced

TABLE 7

Literature Overview on Automatic Skills' Assessment

Technique	Methodology	Scenario	Tasks	Metrics	Sample population	Average success rate (%)	Ref.
HMM	Train expert model	LapSim "Basic Skills"	Touch/coordination	Acceleration (STFT)	4 Novices 2 Experts	100	[107] [108]
	Train mixed model	Box trainer	Touch/ coordination	Trajectory	9 Novices 2 Experts	93 (OSATS)	[109]
	One model/procedure step & skill level	Animal + BlueDragon	Cholecystectomy	Force signatures, idle states	2 × R1, R3, R5 2 Experts	N/A	[49]
	One model/motion signature & skill level	Da Vinci®	Suture	Motion signatures, idle states	1 Novice 1 Resident 1 Expert	58.05	[110] [111]
	One model/motion signature & skill level	Da Vinci®	Suture	Motion signatures	# Experts # Intermediates # Novices (9 surgeons total)	100	[106]
	One model/skill level	Da Vinci®	Suture	Motion signatures	# Experts # Intermediates # Novices (9 surgeons total)	94.7	[106]
MM	One model/skill level	Da Vinci®	Suture	Motion signatures	# Experts # Intermediates # Novices (9 surgeons total)	100	[106]
	One model/subject	Animal + BlueDragon	Suture	Force signatures, idle states	5 × R1, R2, R3, R4, R5 5 Experts	86	[50]
	One model/procedure step & skill level	Animal + BlueDragon	Cholecystectomy/ Nissen	Force signatures, idle states	5 Novices 5 Experts	87.5	[59]
Fuzzy	1 Classifier/skill level & task	MIST-VR core skills	Peg grasp, peg transfer,	Time, errors, economy of movements, score	4 Novices 4 Intermediates 4 Experts	N/A	[114]
	2 Classifiers/task	MIST-VR core skills	Suture	Time, errors, collisions/tissue damage	10 Novices 8 Intermediates 8 Experts	38.25	[115]
LDA	1 Classifier/skill level & task	Box trainer + TrEndo	Peg grasp, peg transfer, bi-manual coordination, cutting	Time, path length, depth, smoothness, angular area, volume	11 Novices 10 Intermediates 10 Experts	74	[46]
	1 Classifier/motion signature	Box trainer	Suture	Motion signatures	2 Experts (15 trials) 1 Intermediate (12 trials)	92	[113]

In [109], success results are reported confronted with OSATS.

HMM = hidden Markov model; MM = Markov model; PCA = principal component analysis; LDA = linear discriminant analysis; # = unknown number.

skills in laboratory settings, (2) training of anatomical protocols and advanced skills with animal models, (3) training advanced procedural skills with tele-surgical applications, and (4) training in the OR.

In the first phases of training, where psychomotor skill acquisition is crucial, VR technologies and tracking in box trainers may be the most suitable choice for the assessment of dexterity [21]. On advanced phases, where cognitive and judgment abilities gain importance, and training gradually moves to the OR, tracking of movements becomes more complicated. As sensor-based tracking systems may interfere in the surgical workflow, we believe the use of endoscopic video information for computer-vision based tracking will be the solution [98]. Moreover, despite current limitations on validation studies, automatic analysis techniques will enhance these systems, allowing to determine not only surgical expertise, but also to unfold and quantify hidden aspects of surgical skill, like the level of automation of tasks (i.e., novices spending much more time in an idle "thinking" state) or the optimal combination of both efficiency and quality metrics for a given task.

Accreditation at each of the formation levels should be achieved before moving on to the next level [5, 8]. However, there is a general lack of consensus on the criteria that should mark official recognition; on the tasks, metrics and assessment methods to employ. Formation and accreditation programs are diverse, and their scope is limited at the topmost to national levels, such as the Fundamentals of Laparoscopic Surgery (FLS), (USA) [119], or the Cobra-alpha courses (The Netherlands) [120].

Specific issues surrounding the change of paradigm in surgical formation will be discussed following the structure proposed in this review [21], attending to (1) the clinical, (2) the technological, and (3) the analytical sides of the problem.

The clinical side deals with the basic question of what needs to be measured and under what conditions; i.e., the tasks and metrics necessary and the training environment where skills will be tested. The number and nature of these tasks are not predefined, but should cover the range of basic psychomotor skills. Concerning metrics, traditional assessment has limited evaluation parameters to: completion time, procedural performance (accuracy, errors), and end-product analysis. Some programs, such as FLS, state that these are sufficient for evaluation purposes [119, 121]. However, to better understand surgical gestures and to exploit the possibilities provided by tracking technologies, a whole range of new efficiency metrics should be carefully considered and analyzed.

Identification of valid metrics can be a difficult process, especially considering their degree of dependence to a given exercise [16]. Error-related metrics, for

FIG. 3. Hierarchical levels of the training pyramid followed in the Minimally Invasive Surgery Centre Jesús Usón (Cáceres, Spain). Author: Minimally Invasive Surgery Centre Jesús Usón.

example, will be closely associated to a task's goal. Moreover, different validation studies show different conclusions for the same metric. Often this is due to the nature and difficulty of the task associated to it [57]. Some of them, however, seem recurrent for different tasks and skills, such as time and path length. Relationship between metrics is also an important factor when considering validation. Examples of this are time-related metrics: speed of task completion is not a useful indicator of skill if the trainee has performed too many mistakes [12]. Finally, when considering several metrics and their nature, it must be taken into account that the means for registering them may vary. For example, motion analysis or path length may be derived for a box trainer task from a precise tracking device; but if information about mistakes performed is desired, it will be necessary to supervise directly or by reviewing the video of the performance [17]. Interrater reliability must be carefully analyzed in those cases [23].

The technological side deals with the capabilities and performance of the systems that will be used in the accreditation process. Traditionally, these have relied on direct or indirect (*via* video review) supervision, and on the usage of structured reports. The challenge nowadays is to find out whether the new automatic tracking and software solutions for HMT and VR can fit within this scheme; and if traditional clinical validation processes can assimilate these systems. In general, VR simulators are nowadays considered as useful tools for skill training, especially in the early stages of formation, during which the resident must acquire and train psychomotor abilities [62]. Thus, it is safe to assume that these systems are a valid supplementary method for surgical training, as effective as that provided by video-based box trainers. There is, however, more reluctance when considering laparoscopic VR simulators as accreditation tools [57, 122]. Their cost and lack of realism, amongst other limitations, hold sway amongst many clinicians. Despite that, validation studies prove that some of the simulators, e.g., MIST-VR or LapSim, are fully capable of addressing formative assessment [18, 123]. Moreover, they are capable of performing unsupervised assessment of trainees in a whole range of efficiency and quality parameters. Nevertheless, these studies are in many cases scarce, and others report negative [124–126] or inconclusive results [77]. This is often the case considering the high variability between studies, concerning participants, definition of surgical levels, and in general the experiments' contour conditions (previous training in the system, number of repetitions allowed, etc.). Further validation studies are thus required to fully prove the potential of VR simulators, especially those regarding predictive skill transfer [57].

HMT tracking devices are cheaper alternatives for VR simulators. In general, HMT's capabilities are generally limited to the acquisition of efficiency metrics. Quality measures are, however, more complicated in these systems, usually requiring the presence of an evaluator, thus compromising immediate feedback for the trainee [28, 53]. Moreover, as Table 6 shows, validation studies are still insufficient and diverse in their nature, and additional proof of their effectiveness for assessment and accreditation purposes is necessary. Besides that, incorporated active trackers tend to limit them to laboratory settings. Therefore, their usage becomes restricted to analysis of box trainer tasks. Passive tracking solutions, such as video analysis, may help boost use of the HMTs for motion analysis on more demanding scenarios (such as the OR). However, research is still on preliminary phases, having been tested on box trainers, and full validation studies must be carried out to ascertain their potential.

The analytical side deals with the evaluation process *per se*. The needs and requirements of the new formation programs require that trainees receive immediate and objective formative feedback of their performance, preferably in the form of comprehensible messages and suggestions that will allow them to improve their skills [104]. Determination of a certification score for summative feedback is a difficult task, partly due to the high number and diverse nature of the metrics available, and the lack of gold standards of surgical competence [16]. The inclusion of advanced automatic evaluation systems, currently an expanding research field, may allow for unsupervised assessment. Nevertheless, studies on this subject are yet few and limited in relation to the number of participants and tasks considered. Moreover, successful classification rates are, in some cases, very low. Additionally, they are conditioned by a training phase, which may be mostly meaningful in standardized tasks (in box trainers), but which might be less valid for clinical procedures. New proof is needed before these systems can be considered for their clinical implantation.

Despite the need for objectivity in new formation programs, the fact remains that, ultimately, final expertise accreditation should come by the hand of an expert mentor. Whilst measurable parameters offer reliable and unbiased data on psychomotor skills, there are many other abilities, such as reaction time, mentality, patient care, handling of stress, or group working capability, which are equally important but more difficult to quantify. Consequently, we consider there is a human, more personal component to the determination of a surgeon's readiness, which will imply evaluation of the aforementioned abilities. These are all important factors that add a human dimension to the qualification process and, consequently, should always be considered.

ACKNOWLEDGMENTS

This article and all related research are being partially funded by the FPU Program of the Spanish Ministry of Science and Innovation.

REFERENCES

1. Cuschieri A. Laparoscopic surgery: Current status, issues and future developments. Surgeon 2005;3:125.
2. Usón J, Sánchez-Margallo FM, Pascual S, Climent S. Formación en Cirugía Laparoscópica Paso a Paso. 4th ed. Cáceres, Spain: Minimally Invasive Surgery Centre Jesús Usón eds, 2010.
3. Breedveld P, Stassen HG, Meijer DW, et al. Observation in laparoscopic surgery: Overview of impending effects and supporting aids. J Laparoendosc Adv 2000;10:231.
4. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. Br J Surg 2004;91:1549.
5. Park AE, Witzke D. The surgical competence conundrum. Surg Endosc 2002;16:555.
6. Ritchie WP. Basic certification in surgery by the American Board of Surgery (ABS). What does it mean? Does it have value? Is it relevant? A personal opinion. Ann Surg 2004; 239:133.
7. Halsted WS. The training of surgeons. Johns Hopkins Hosp Bull 1904;15:267.
8. Satava RM, Cuschieri A, Hamdorf J. Metrics for objective assessment. Surg Endosc 2003;17:220.
9. Sidhu R, Grober E, Musselman L, et al. Assessing competency in surgery: Where to begin? Surgery 2004;135:6.
10. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: From selection to certification. Curr Prob Surg 2002;39:573.
11. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005;190:107.
12. Fried GM, Feldman LS. Objective assessment of technical performance. World J Surg 2008;32:156.
13. Harden RM, Stevenson M, Downie WW, et al. Assessment of clinical competence using objective structured examination. Br Med J 1975;1:447.
14. Martin J, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84.
15. Moorthy K, Munz Y, Sarker SK, et al. Objective assessment of technical skills in surgery. Br Med J 2003;327:1032.
16. Chmarra MK, Grimbergen CA, Jansen FW, et al. How to objectively classify residents based on their psychomotor laparoscopic skills? Minim Invasive Ther Allied Technol 2010;19:2.
17. Dosis A, Aggarwal R, Bello F, et al. Synchronized video and motion analysis for the assessment of procedures in the operating theater. Arch Surg 2005;140:293.
18. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, et al. Objective assessment of technical surgical skills. Br J Surg 2010;97:972.
19. Satava RM. Historical review of surgical simulation—A personal perspective. World J Surg 2008;32:141.
20. Chmarra MK, Grimbergen CA, Dankelman J. Systems for tracking minimally invasive surgical instruments. Minim Invasive Ther Allied Technol 2007;16:328.
21. Oropesa I, Sánchez-González P, Cano AM, et al. Objective evaluation methodology for surgical motor skills assessment. Minim Invasive Ther Allied Technol 2010;10:55.
22. Cotin S, Stylopoulos N, Ottensmeyer M, et al. Metrics for laparoscopic skills trainers: The weakest link!. In: Dohi T, Kikinis R, Eds. Medical image computing and computer-assisted intervention — MICCAI 2002. Berlin/Heidelberg: Springer, 2002:35–43.
23. Lamata P. Methodologies for the analysis, design and evaluation of laparoscopic surgical simulators. PhD Thesis. Louvain: Presses univ de Louvain, 2006.
24. Sánchez-Peralta LF, Sánchez-Margallo FM, Moyano-Cuevas JL, et al. Construct and face validity of SINERGIA laparoscopic virtual reality simulator. Int J Comput Assist Radiol Surg 2010; 5:307.
25. van Sickle KR, McClusky DA 3rd, Gallagher AG, et al. Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surg Endosc 2005;19:1227.
26. Pellen MG, Horgan LF, Barton JR, et al. Construct validity of the ProMIS laparoscopic simulator. Surg Endosc 2009;23:130.
27. Broe D, Ridgway PF, Johnson S, et al. Construct validation of a novel hybrid surgical simulator. Surg Endosc 2006;20:900.
28. Pellen M, Horgan L, Barton JR, et al. Laparoscopic surgical skills assessment: Can simulators replace experts? World J Surg 2009;33:440.
29. Yamaguchi S, Konishi K, Yasunaga T, et al. Construct validity for eye–hand coordination skill on a virtual reality laparoscopic surgical simulator. Surg Endosc 2007;21:2253.
30. Zhang A, Hünerbein M, Dai Y, et al. Construct validity testing of a laparoscopic surgery simulator (Lap Mentor®): Evaluation of surgical skill with a virtual laparoscopic training simulator. Surg Endosc 2008;22:1440.
31. Aggarwal R, Crochet P, Dias A, et al. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. Br J Surg 2009;96:1086.
32. McDougall EM, Corica FA, Boker JR, et al. Construct validity testing of a laparoscopic surgical simulator. J Am Coll Surg 2006;202:779.
33. Verdaasdonk EGG, Stassen LPS, Schijven MP, et al. Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. Surg Endosc 2007; 21:1406.
34. Eriksen JR, Grantcharov T. Objective assessment of laparoscopic skills using a virtual reality stimulator. Surg Endosc 2005;19:1216.
35. Sherman V, Feldman L, Stanbridge D, et al. Assessing the learning curve for the acquisition of laparoscopic skills on a virtual reality simulator. Surg Endosc 2005;19:678.
36. Duffy A, Hogle N, McCarthy H, et al. Construct validity for the LAPSIM laparoscopic surgical simulator. Surg Endosc 2005; 19:401.
37. Larsen CR, Grantcharov T, Aggarwal R, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc 2006;20:1460.
38. Woodrum DT, Andreatta PB, Yellamanchilli RK, et al. Construct validity of the LapSim laparoscopic surgical simulator. Am J Surg 2006;191:28.
39. Schreuder HWR, van Dongen KW, Roeleveld SJ, et al. Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. Obstet Gynecol 2009;200:540.
40. van Dongen KW, Tournoij E, van der Zee DC, et al. Construct validity of the LapSim: Can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 2007;21:1413.
41. Panait L, Bell RL, Roberts KE, et al. Designing and validating a customized vitual reality-based laparoscopic skills curriculum. J Sur Educ 2008;65:413.
42. Gallagher A, Lederman A, McGlade K, et al. Discriminative validity of the minimally invasive surgical trainer in virtual reality (MIST-VR) using criteria levels based on expert performance. Surg Endosc 2004;18:660.
43. McNatt SS, Smith CD. A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. Surg Endosc 2001;15:1085.
44. Maithel S, Sierra R, Korndorffer J, et al. Construct and face validity of MIST-VR, Endotower, and CELTS. Surg Endosc 2006; 20:104.
45. Stylopoulos N, Cotin S, Maithel S, et al. Computer-enhanced laparoscopic training system (CELTS): Bridging the gap. Surg Endosc 2004;18:782.

46. Chmarra MK, Klein S, de Winter JCF, et al. Objective classification of residents based on their psychomotor laparoscopic skills. Surg Endosc 2010;24:1031.
47. Egi H, Okajima M, Yoshimitsu M, et al. Objective assessment of endoscopic surgical skills by analyzing direction-dependent dexterity using the Hiroshima University Endoscopic Surgical Assessment Device (HUESAD). Surg Today 2008;38:705.
48. Sokollik C, Gross J, Buess G. New model for skills assessment and training progress in minimally invasive surgery. Surg Endosc 2004;18:495.
49. Rosen J, Solazzo M, Hannaford B, et al. Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. Comput Aided Surg 2002;7:49.
50. Rosen J, Brown JD, Chang L, et al. Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. IEEE Trans Biomed Eng 2006;53:399.
51. Bann SD, Khan MS, Darzi AW. Measurement of surgical dexterity using motion analysis of simple bench tasks. World J Surg 2003;27:390.
52. Datta V, Mackay S, Mandalia M, et al. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg 2001; 193:479.
53. Moorthy K, Munz Y, Dosis A, et al. Bimodal assessment of laparoscopic suturing skills. Surg Endosc 2004;18:1608.
54. Brydges R, Classen R, Larmer J, et al. Computer-assisted assessment of one-handed knot tying skills performed within various contexts: A construct validity study. Am J Surg 2006;192:109.
55. Francis NK, Hanna GB, Cuschieri A. The performance of master surgeons on the advanced dundee endoscopic psychomotor tester contrast validity study. Arch Surg 2002;37:841.
56. Horeman T, Rodrigues SP, Jansen FW, et al. Force measurement platform for training and assessment of laparoscopic skills. Surg Endosc 2010;24:3102.
57. Thijssen AS, Schijven MP. Contemporary virtual reality laparoscopy simulators: Quicksand or solid grounds for assessing surgical trainees? Am J Surg 2010;199:529.
58. Chmarra MK, Jansen FW, Grimbergen CA, et al. Retracting and seeking movements during laparoscopic goal-oriented movements. Is the shortest path length optimal? Surg Endosc 2008;22:943.
59. Rosen J, Hannaford B, Richards C, et al. Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. IEEE Trans Biomed Eng 2001;48:579.
60. Cuschieri A, Francis SN, Crosby J, et al. What do master surgeons think of surgical competence and revalidation? Am J Surg 2001;182:110.
61. Dankelman J, Chmarra MK, Verdaasdonk EGG, et al. Fundamental aspects of learning minimally invasive surgical skills. Minim Invasive Ther Allied Technol 2005;14:247.
62. Gurusamy KS, Aggarwal R, Palanivelu L, et al. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2009;1:CD006575.
63. Dunkin B, Adrales GL, Apelgren K, et al. Surgical simulation: A current review. Surg Endosc 2007;21:357.
64. Neary PC, Boyle E, Delaney CP, et al. Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; Results from the first course for experienced senior laparoscopic surgeons. Surg Endosc 2008; 22:2301.
65. Ahlberg G, Kruuna O, Leijonmarck CE, et al. Is the learning curve for laparoscopic fundoplication determined by the teacher or the pupil? Am J Surg 2005;189:184.
66. Madan AK, Frantzides CT, Sasso LM. Laparoscopic baseline ability assessment by virtual reality. Laparoendosc J Surg Tech A 2005;15:13.
67. Stefanidis D, Korndorffer JR Jr, Black FW, et al. Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. Surgery 2006;140:252.
68. Grantcharov TP, Rosenberg J, Pahle E, et al. Virtual reality computer simulation. Surg Endosc 2001;15:242.
69. Okrainec A, Greco E, Aarts MA, et al. Virtual reality training on basic laparoscopic tasks *versus* virtual reality training of an entire surgical procedure: A randomized controlled trial using real world operations as an outcome. Proceedings of the 2008 Annual Meeting of the Association for Surgical Education, Toronto, Canada, 2008.
70. Greco E, Escallon J, Grantcharov T, et al. Do performance measures on the LapMentor VR simulator predict FLS performance? Can J Surg 2008;51:15.
71. Youngblood PL, Srivastava S, Curet M, et al. Comparison of training on two laparoscopic simulators and assessment of skills transfer to surgical performance. J Am Coll Surg 2005; 200:546.
72. Newmark J, Dandolu V, Milner R, et al. Correlating virtual reality and box trainer tasks in the assessment of laparoscopic surgical skills. Obstet Gynecol 2007;197:546.
73. Hassan I, Osei-Agymang T, Radu D, et al. Simulation of laparoscopic surgery–Four years' experience at the Department of Surgery of the University Hospital Marburg. Wien Klin Wochenschr 2008;120:70.
74. Ritter EM, Kindelan TW, Michael C, et al. Concurrent validity of augmented reality metrics applied to the fundamentals of laparoscopic surgery (FLS). Surg Endosc 2007;21:1441.
75. Botden SM, Buzink SN, Schijven MP, et al. Augmented versus virtual reality laparoscopic simulation: What is the difference? A comparison of the ProMIS augmented reality simulator versus LapSim virtual reality laparoscopic simulator. World J Surg 2007;31:764.
76. McCluney A, Feldman L, Fried G. Validation of the ProMIS hybrid simulator using a standard set of laparoscopic tasks. Proc Sages 2006.
77. Verdaasdonk EGG, Stassen LPS, Monteny LJ, et al. Validation of a new basic virtual reality simulator for training of basic endoscopic skill: The SIMENDO. Surg Endosc 2006;20:511.
78. Lamata P, Gomez EJ, Bello F, et al. Conceptual framework for laparoscopic VR simulators. IEEE Comput Graph Appl 2006; 26:69.
79. Chmarra MK, Dankelman J, van den Dobbelsteen JJ, et al. Force feedback and basic laparoscopic skills. Surg Endosc 2008;22:2140.
80. Panait L, Akkary E, Bell RL, et al. The role of haptic feedback in laparoscopic simulation training. J Surg Res 2009;156:312.
81. Issenberg SB, Mcgaghie WC, Petrusa ER, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. Med Teach 2005;27:10.
82. Lynch J, Aughwane P, Hammond TM. Video games and surgical ability: A literature review. J Surg Educ 2010; 67:184.
83. Lamata P, Gómez EJ, Sánchez-Margallo FM, et al. SINERGIA laparoscopic virtual reality simulator: Didactic design and technical development. Comput Methods Programs Biomed 2007;85:273.
84. Lamata P, Gomez EJ, Hernandez FL, et al. Understanding perceptual boundaries in laparoscopic surgery. IEEE Trans Bio Eng 2008;55:866.
85. Sánchez-González P, Oropesa I, Sánchez-Margallo FM, et al. A virtual reality simulator-based system for MIS objective evaluation. Minim Invasive Ther Allied Technol 2010;10:15.
86. Hanna G, Frank T, Cuschieri A. Objective assessment of endoscopic knot quality. Am J Surg 1997;174:410.
87. Solis J, Oshima N, Ishii H, et al. Towards understanding the suture/ligature skills during the training process using WKS-2RII. Int J Comput Assist Radiol Surg 2008;3:231.

88. Sánchez-Margallo FM, Pagador JB, Sánchez-Peralta LF, et al. Estudios de validez de un dispositivo háptico de realidad aumentada (ARH) para la evaluación objetiva de la sutura laparoscópica intracorpórea. Cir Esp 2010;88:218.
89. Datta V, Mandalia M, Mackay S, et al. Relationship between skill and outcome in the laboratory-based model. Surgery 2002;131:318.
90. Datta V, Bann S, Mandalia M, et al. The surgical efficiency score: A feasible, reliable, and valid method of skills assessment. Am J Surg 2006;192:372.
91. Macmillan AIM, Cuschieri A. Assessment of innate ability and skills for endoscopic manipulations by the Advanced Dundee Endoscopic Psychomotor Tester: Predictive and concurrent validity. Am J Surg 1999;177:274.
92. Hanna G, Drew T, Clinch P, et al. Computer-controlled endoscopic performance assessment system. Surg Endosc 1998; 12:997.
93. Chmarra MK, Bakker NH, Grimbergen CA, et al. TrEndo, a device for tracking minimally invasive surgical instruments in training setups. Sensor Actuat: A Physical 2006;126:328.
94. Pagador JB, Sánchez LF, Sánchez JA, et al. Augmented reality haptic (ARH): An approach of electromagnetic tracking in minimally invasive surgery. Int J Comput Assist Radiol Surg 2011; 6:257.
95. Rosen J, Brown JD, Barreca M, et al. The Blue DRAGON–A system for monitoring the kinematics and the dynamics of endoscopic tools in minimally invasive surgery for objective laparoscopic skill assessment. Stud Health Technol Inform 2002; 85:412.
96. Hayter MA, Friedman Z, Bould MD, et al. Validation of the Imperial College Surgical Assessment Device (ICSAD) for labour epidural placement. Can J Anesth 2009;56:49.
97. Aggarwal R, Grantcharov T, Moorthy K, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg 2007;245:992.
98. Sánchez-González P, Cano AM, Oropesa I, et al. Laparoscopic video analysis for training and image guided surgery. Minim Invasive Ther Allied Technol 2010 (in press).
99. Voros S, Long J, Cinquin P. Automatic localization of laparoscopic instruments for the visual serving of an endoscopic camera holder. Med Image Comput Comput Assist Interv 2006;9:535.
100. Doignon C, Nageotte F, Maurin B, et al. Pose estimation and feature tracking for robot assisted surgery with medical imaging. In: Kragic D, Kyrik V, Eds. Unifying perspectives in computational and robot vision, vol. 8. New York: Springer US, 2008:79.
101. Sánchez-Margallo JA, Sánchez-Margallo FM, Pagador JB, et al. Video-based assistance system for training in minimally invasive surgery. Minim Invasive Ther Allied Technol 2010 (In press).
102. Pagador JB, Sánchez-Margallo FM, Sánchez-Peralta LF, et al. Objective assessment of basic laparoscopic skills using automatic video-based technique. Minim Invasive Ther Allied Technol 2010;19:55.
103. Cano A, Gayá F, Lamata P, et al. Laparoscopic tool tracking method for augmented reality surgical applications. In: Bello F, Edwards P, Eds. Biomedical Simulation. Berlin/Heidelberg: Springer, 2008:191–6.
104. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. Ann Surg 2004;3:518.
105. Jain AK, Duin RPW, Mao J. Statistical pattern recognition: A review. IEEE Trans Pattern Anal Machine Intell 2000;22:4.
106. Reiley CE, Hager GD. Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. Med Image Comput Comput Assist Interv 2009;12:435.
107. Megali G, Sinigaglia S, Tonet O, et al. Modelling and evaluation of surgical performance using hidden Markov models. IEEE Trans Biomed Eng 2006;53:1911.
108. Megali G, Sinigaglia S, Tonet O, et al. Understanding expertise in surgical gesture by means of hidden Markov models. Proc IEEE RAS EMBS Int Conf Biomed Robot Biomechatron 2006:625.
109. Leong JJH, Nicolaou M, Atallah L, et al. HMM assessment of quality of movement trajectory in laparoscopic surgery. Comput Aided Surg 2007;12:335.
110. Murphy TE, Vignes CM, Yuh DD, et al. Automatic motion recognition and skill evaluation for dynamic tasks. Proc Euro Haptics 2003:363.
111. Murphy TE. Towards objective surgical skill evaluation with hidden Markov model-based motion recognition. Baltimore: Johns Hopkins University, 2004.
112. Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989;77:257.
113. Lin HC, Shafran I, Yuh D, et al. Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. Comput Aided Surg 2006;11:220.
114. Huang J, Payandeh S, Doris P, et al. Fuzzy classification: Towards evaluating performance on a surgical simulator. Stud Health Technol Inform 2005;111:194.
115. Hajshirmohammadi I, Payandeh S. Fuzzy set theory for performance evaluation in a surgical simulator. Presence 2007; 16:603.
116. Estebanez B, Jimenez G, Muñoz VF, et al. Minimally invasive surgery maneuver recognition based on surgeon model. Intelligent Robots and Systems, 2009 IROS 2009 IEEE/RSJ International Conference on, 2009;5522.
117. Blum T, Padoy N, Feußner H, et al. Modeling and online recognition of surgical phases using hidden Markov models. In: Metaxas D, Axel L, Fichtinger G, Székely G, Eds. Medical image computing and computer-assisted intervention–MICCAI 2008. Berlin/Heidelberg: Springer, 2008:627.
118. Kneebone R. Simulation in surgical training: Educational issues and practical implications. Med Educ 2003;37:267.
119. Ritter EM, Scott D. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. Surg Innov 2007;14:107.
120. Hiemstra E, Kolkman W, Jansen FW. Skills training in minimally invasive surgery in Dutch obstetrics and gynecology residency curriculum. Gynecol Surg 2008;5:321.
121. Scott DJ, Ritter EM, Tesfay ST, et al. Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training. Surg Endosc 2008;8:1887.
122. Feldman LS, Sherman V, Fried GM. Using simulators to assess laparoscopic competence: Ready for widespread use? Surgery 2004;135:28.
123. Oropesa I, Lamata P, Sánchez-González P, et al. Virtual reality simulators for objective evaluation on laparoscopic surgery: Current trends and benefits. In: Kim J, Ed. Virtual reality. Rijeka: Intech Open Access Publisher, 2011.
124. Langelotz C, Kilian M, Paul C, et al. LapSim virtual reality laparoscopic simulator reflects clinical experience in German surgeons. Langenbecks Arch Surg 2005;390:534.
125. Ro CY, Toumpoulis IK, Ashton RC Jr, et al. The LapSim: A learning environment for both experts and novices. Stud Health Technol Inform 2005;111:414.
126. van Sickle KR, Ritter EM, McClusky DA III, et al. Attempted establishment of proficiency levels for laparoscopic performance on a national scale using simulation: The results from the 2004 SAGES minimally Invasive Surgical Trainer-Virtual Reality (MIST-VR) learning center study. Surg Endosc 2007;21:5.