Skill-oriented and Performance-driven Adaptive Curricula for Training in Robot-Assisted Surgery using Simulators: a Feasibility Study

Andrea Mariani[1,2], *Student Member, IEEE,* Edoardo Pellegrini[2], and Elena De Momi[2], *Member, IEEE*

***Abstract* — *Objective:* Virtual Reality (VR) simulators represent a remarkable educational opportunity in order to acquire and refine surgical practical skills. Nevertheless, there exists no consensus regarding a standard curriculum of simulation-based training. This study introduces an automatic, adaptive curriculum where the training session is real-time scheduled on the basis of the trainee's performances.**

***Methods:* An experimental study using the master console of the da Vinci Research Kit (Intuitive Surgical Inc., Sunnyvale, US) was carried out to test this approach. Tasks involving fundamental skills of robotic surgery were designed and simulated in VR. Twelve participants without medical background along with twelve medical residents were randomly and equally divided into two groups: a control group, self-managing the training session, and an experimental group, undergoing the proposed adaptive training.**

***Results:* The performances of the experimental users were significantly better with respect to the ones of the control group after training (non-medical: $p<0.01$; medical: $p=0.02$). This trend was analogous in the non-medical and medical populations and no significant difference was identified between these two classes (even in the baseline assessment).**

***Conclusion:* The analysis of the learning of the involved surgical skills highlighted how the proposed adaptive training managed to better identify and compensate for the trainee's gaps. The absence of initial difference between the non-medical and medical users underlines that robotic surgical devices require specific training before clinical practice.**

***Significance:* This feasibility study could pave the way towards the improvement of simulation-based training curricula.**

***Index Terms* — Virtual Reality Simulators, Training and Motor Learning, Robot-Assisted Surgery, Adaptive Logics, Skill Assessment**

I. Introduction

Simulation-based education for the development of practical skills has come to the forefront in the recent years. This growth has affected different fields including medical and, in particular, surgical education [1]. Simulation can be defined as the creation of interactive environments that replicate a real-world scenario, although not identical to "real life" [2]. This has paved the way to new learning approaches that could replace the classical Halstedian paradigm (*see one do one, teach one*). In fact, a direct ethical advantage of simulation-based training in surgery is preventing patients from being subjected to any risk or complication during the learning phase [3]. In other words, a preliminary hands-on experience can be gathered before operating on real patients. Additionally, simulators are characterized by an easier access with respect to time-consuming and expensive practice in the operating room. This allows to overcome the constraints associated to a limited case volume for surgical training [4]. This is all the more important since the increasing deployment of Robot-Assisted Surgery (RAS) has introduced the need of learning also how to deal with new interfaces between the surgeon and the surgical instruments. In fact, this approach has become currently adopted in a wide area of surgical interventions including urology, gynecology and general abdominal surgery [5][6]. This technique is able to combine the advantages of Minimally Invasive Surgery (MIS) with respect to open surgery [7] and the benefits of robotic surgery compared to standard MIS [8]. Nonetheless, the efficacy of this approach is strictly related to the surgeon's experience [9]. In fact, this technique implies new control modalities, where the user has to manipulate a couple of robotic masters in order to position the surgical instruments and the endoscopic camera inside the patient. This teleoperative approach detrimentally involves the need of learning how to deal with the new dynamics of the manipulators [10] and how to compensate for the absence of haptic feedback [11]. Together with the complexity of surgery as a sensorimotor task, these new technologies lead to a distinct necessity of efficient training modalities. A recent survey-based study of residents enrolled in general surgery training program revealed that 78% of senior general surgery residents have been involved in a robot-assisted case, yet more than half of these residents have not received any formal training to this platform [12]. Simulators can overcome these issues and help in allowing a safe introduction of robot-

[1]Andrea Mariani is with the BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa 56121, Italy (e-mail: andrea.mariani@santannapisa.it).

[2]Andrea Mariani, Edoardo Pellegrini and Elena De Momi are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133, Italy.

based technologies in the surgical scenario.

In order to maximize the learning benefit, there is a clear need for structured training curricula [13]. Three essential phases compose a general structured training program in robotic surgery [14]: the *preclinical phase* (didactic and skill training, comparison of surgical robotic methods with prior training and clinical experiences, and understanding of the robot technology); secondly, the *bedside assistant phase* (working as co-surgeon, learning trocar and robot placement, instrumentation, troubleshooting); and finally, the *operative console phase* (performing parts of the robotic operation). Simulators can be an effective training method in the initial preclinical stage by allowing a first approach to the surgical platform and a practical skill development [15-16]. This means a simulation-based training platform has to transfer the required surgical skills as efficiently as possible. Exercises should have *face validity* (i.e. the simulation resembles the real task), *content validity* (i.e. the intended content domain is measured by the assessment exercise), *construct validity* (i.e. the ability to differentiate between groups with different levels of competence), and *predictive validity* (i.e. capability to predict future performances and retain the acquired skills) [17]. For instance, the dV-Trainer by Mimic Technologies Inc. (Seattle, WA) is a virtual reality simulator that replicates the master console of the da Vinci robot by Intuitive Surgical Inc. (Sunnyvale, CA). Several studies have tested the features previously listed and the transferability of the skills acquired by training with this simulator to the real surgical field [18-19]. Processing the concept of *construct validity*, the validation of simulators should now follow Messick's framework, where five distinct sources of validity evidence are investigated to build a validity argument [20]. These sources involve content, response process, inner structure, relation with other variables, and consequences of the assessment test. In [21], 33 studies using Messick's validity framework to validate surgical simulators are reviewed.

Another key aspect while designing a training platform is the quantification of the user's performance, which should be carried out as soon and objectively as possible [22-23]. The assessment of technical skills by observation, as it typically occurs in the operating room, can be time-expensive, subjective and then lacking of reliability [23-24]. Simulators cover this limitation by adding the possibility of quantitative measurements of the trainee's performance by computing intra-task metrics [25-27]. A metric can be defined as a variable that is sampled during the task execution. It can be task-independent (e.g. time, number of movements, path length of the end effector, as in [28]) or task-dependent (e.g. maximum cutting velocity, number of cutting segments, as in [29]). The constant monitoring of these metrics, as well as kinematics data associated to the user's interaction with the masters, can provide accurate and objective measures about the trainee's learning.

Simulators have shifted the paradigm of a mentor-guided learning towards a self-directed training, with the consequent reduction of program costs [30]. Nevertheless, the success of this approach relies on the ability of the trainee to correctly perform self-assessment and management. Especially in the early stages of training, this condition is not totally accomplished [31]. Some recent studies have investigated the

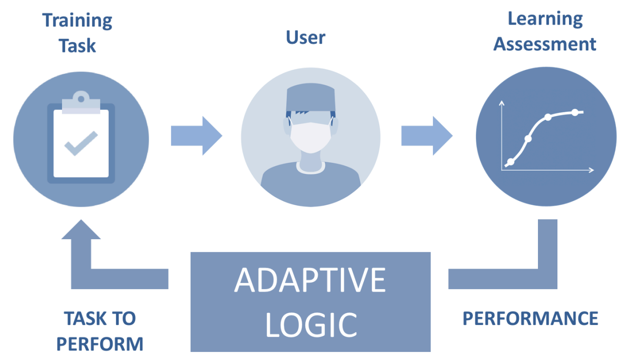

Figure 1 *Performance-based adaptive curriculum for robotic surgical training using simulators: the trainee performs a certain task and undergoes an assessment of his/her learning that is quantified in a performance. This last (together with the previous ones, if available) is employed in order to decide which exercise performing next.*

possibility to apply computer-based feedback during training. In [32], Kowalewski et al. proposed a training system for laparoscopy based on sensor and expert models. This system is able to generate automated real-time feedback to optimize the trainee's learning. Malpani et al. developed an automated coach that provides real-time cues (like visual overlays and video demonstrations) for robotic surgical training and they demonstrated its feasibility on 16 subjects [33]. Out of the surgical field, Rauter et al. studied training by means of a robotic rowing simulator and they introduced an automated selection of the feedback type to incorporate human expertise in the learning process [34]. Enayati et al. addressed this topic by analyzing the application of an *assistance-as-needed* haptic feedback during training: this guidance was able to optimize the learning of some performance metrics (such as time) in the field of robotic surgery [35]. In this work, we address the modulation of the training curriculum. In fact, while training with simulators, there exists the need for defining a training session curriculum (i.e. which exercises perform and in which order). In order to overcome the restrictions of a static schedule, adaptive approaches have been proposed in other educational fields [36]. Adaptive training can be defined as a training modality in which the task is varied as a function of how well the trainee performs [37]. Such an approach features three key elements: a performance measurement module, an adjustable task feature (the adaptive variable) and an adaptive logic that automatically changes the adaptive variable as a function of the performance measurement. This adaptivity should also increase the training efficiency since effective learning takes place only when training occurs at an appropriate level of challenge for the trainee [38].

Our study proposes an adaptive training curriculum for robotic surgery using simulators that automatically schedules the training session on the basis of an objective assessment of the trainee's performances. This closed-loop framework is depicted in Figure 1. The research question can be summarized as follows: *can an adaptive training achieve better final performances compared to a self-managed training, where the word adaptive refers to the fact that the training modules are arranged in an automatic and performance–based way*? The

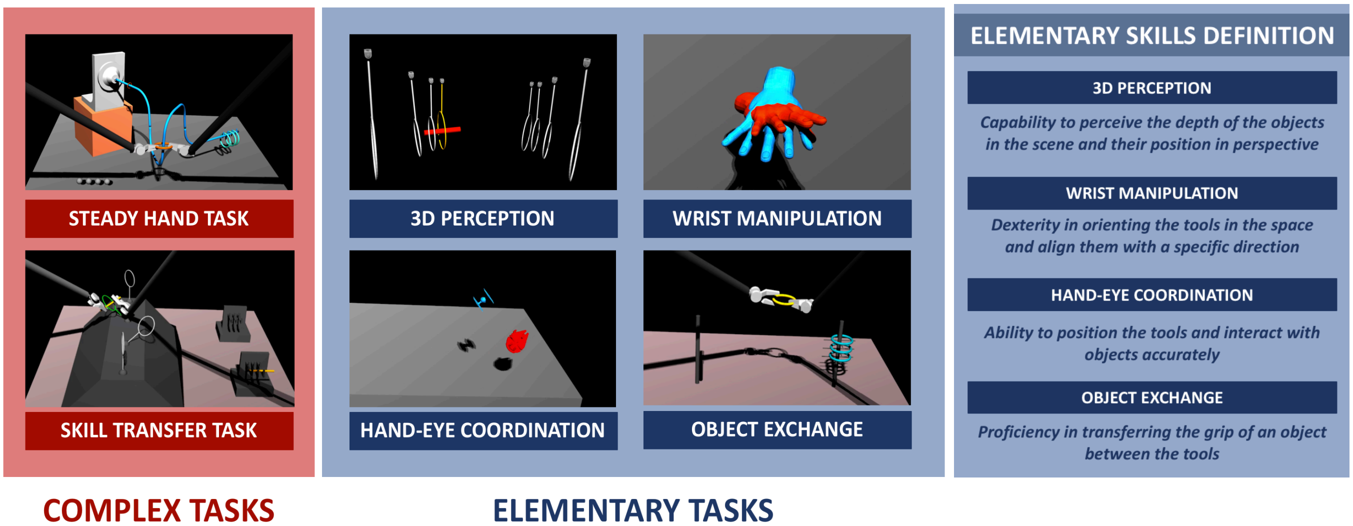

Figure 2 *The task pool is composed by complex and elementary tasks. The former contains all the elementary skills that the latter individually focuses on.*

research hypothesis is tested by analyzing both participants without medical background and a population of medical residents. This work represents an extension of a preliminary pilot study of ours [39].

II. Methods

In order to test the research hypothesis, we exploited a virtual reality simulation environment that allowed us to design skill-oriented tasks and to objectively quantify the trainee's performances. The experimental study was designed primarily to assess the efficacy of the adaptive training on the final performances after training, and secondly to analyze the skill learning and the relevance of previous medical experience. The following paragraphs describe the experimental setup and the acquisition protocol, as well as the specifications of the adaptive training logic.

A. Experimental Setup

The robotic platform employed in this study was the surgeon console of a da Vinci Research Kit (dVRK). This is a standard da Vinci surgical system (Intuitive Surgical, CA, USA) that is integrated with control hardware and software in order to directly access the measurements relative to the joints of each manipulator of the system [40]. The dVRK's surgeon console includes a foot-pedal tray, a stereo viewer (each viewer has a resolution of 640x480 and refresh rate of 59.94 Hz) and two master manipulators (each one equipped with 7 actuated joints and a passive gripper).

The VR environment was designed using our recently developed Assisted Teleoperation with Augmented Reality (ATAR) framework [41]. The developed software architecture for this study encompasses four Robot Operating System (ROS) nodes, namely: core simulation, teleoperation, Graphical User Interface (GUI) and bridge to dVRK controllers. The core simulation node generates the graphics and physics of the virtual objects, along with the task logic. The simulated environment runs at 25 Hz. The 3D graphics are produced using the Visualization Toolkit (VTK) and OpenGL libraries and are sent to the dVRK stereo displays through two output ports of a GeForce GTX 980 Ti GPU (Nvidia Corp.) with a refresh rate of 25 Hz. The Bullet physics library [42] was deployed to perform the dynamic simulation of 3D virtual objects. The robotic tools are simulated as kinematic objects resembling the da Vinci endoscopic tools and their pose is constrained to that of the master devices through the teleoperation node and with a translation down-scaling factor of 2. To ensure real-time interactivity with realistic collision detection, mesh objects are approximated through convex hull decomposition [43]. Through the GUI node, the operator can select the task to perform, as well as controlling the record of the current session data (which is performed at 30 Hz).

B. Training Tasks

As initially stated, simulators are involved in technical skills learning: we developed a curriculum composed by elementary and complex tasks. The former aims at training a single fundamental skill of robotic surgery, while the latter involves multiple skills at once. Starting from a complex bimanual visuo-motor task (analyzed in a previous study of ours [35]), we identified the involved elementary skills. In particular, this choice was guided by the analysis of a *skill deconstruction list* [34] generated by robotic operations observation and by interviews with clinical experts. These skills are listed and defined in Figure 2. Notice that, although the robotic platform enables 3D vision by a stereoscopic viewer, the trainee has to learn how to deal with this new vision modality [45-46]. We also highlight that all the elementary tasks involve a single arm at once, except for the object exchange: this latter task is meant to be elementary since it trains and evaluates just the grip transfer from an instrument to the other. Additionally, we designed a second complex task that involves the same basic skills of the former. This task does not make part of the training curriculum, but it appears just in the final evaluation stage in order to test the transferability of the acquired skills on a task which is totally new to the trainee. Figure 2 contains a picture of each virtual reality exercise, whose logic is explained in the following lines.

1. **Steady Hand Task** – *Complex*
   The user moves a ring along a curved wire pathway, while attempting to avoid the ring and wire making contact and keeping the ring's plane perpendicular to the wire's tangent. In each repetition, the subject has to grasp a ring from the right side of the wire with the right hand tool, carry it to the middle of the wire and transfer the ring to the left tool and finally carry the ring to the left end of the wire.
2. **Skill Transfer Task** – *Complex*
   The user grabs a needle and performs an instrument to instrument exchange through randomly oriented rings. While conducting the transfer, he/she has to keep the needle orthogonal with respect to the ring's plane and passing through the ring's center.
3. **3D Perception Task** – *Elementary*
   The user controls a cylinder that moves in the space. The purpose is to insert the cylinder sequentially into rings placed at different depth levels.
4. **Wrist Manipulation Task** – *Elementary*
   The user has the control of the orientation of one virtual hand and he/she has to rotate his wrist in order to keep it overlapped to a second, concentric and randomly rotating hand.
5. **Hand-Eye Coordination Task** – *Elementary*
   The user controls the position of a pointer and tries to keep it as close as possible to a target that moves along a random path in the space.
6. **Object Exchange Task** – *Elementary*
   The user has to grasp a ring from the right tower and to place it on the target tower after performing an instrument-to-instrument exchange.

C. Performance Assessment

The performances were objectively assessed on the basis of five task metrics (Table I). These metrics have shown to be capable of building learning curves across training in [35]. All the metrics were expressed as values ranging from 0 to 1. To achieve this goal, a metric m was normalized and saturated as:

$$m^* = sat(0, \frac{m_{best} - m}{m_{best} - m_{worst}}, 1) \quad (1)$$

where $sat(a, x, b)$ saturates the input x between the minimum a and the maximum b. The ideal value of each metric m_{best} was determined by taking into consideration the performances of three surgeons (all with number of robotic surgical cases > 300 so that they could be regarded as experts), while the lower boundary m_{worst} was tuned according to novice users' data. Precisely, m_{best} and m_{worst} were computed as the average of the 5 best and worst values of each metric ever recorded, considering 50 repetitions of each task (performed by expert and novice users, as previously stated). Finally, the performance P_i in the ith elementary task was computed as a weighted sum of the associated metrics:

$$P_i = \frac{\sum \beta_j m_j}{\sum \beta_j} \in [0,1] \quad (2)$$

where β_j is the weight associated to the jth metric m_j in the ith elementary task and j ranges from 1 to 5.

Table I *Task Metrics*

INDEX	NAME	DESCRIPTION	UNIT
m_1	DEPTH ERROR*	Distance between tool and target along the perspective direction	*mm*
m_2	TIME	Time needed to complete the task	*sec*
m_3	POSITION ERROR*	Linear distance between the actual and the ideal position of the manipulator	*mm*
m_4	ORIENTATION ERROR*	Angular difference between the actual and ideal pose of the manipulator	*rad*
m_5	EXCHANGE ERROR	Number of drops while performing an object transfer	/

METRIC WEIGHTS		m_1	m_2	m_3	m_4	m_5
COMPLEX TASKS	DEPTH PERCEPTION	7	1	1	1	0
	HAND-EYE COORDINATION	1	2	0	6	1
	WRIST ARTICULATION	1	0	9	0	0
	OBJECT EXCHANGE	0	3	0	0	7
ELEMENATRY TASKS	DEPTH PERCEPTION	7	2	0	1	0
	HAND-EYE COORDINATION	1	2	0	7	0
	WRIST ARTICULATION	2	0	8	0	0
	OBJECT EXCHANGE	2	1	0	0	7

* These metrics are computed as average value in each task repetition

Table II *Participants Statistics*

		CONTROL GROUP	EXPERIMENTAL GROUP
NON - MEDICAL PARTICIPANTS			
	SEX	2 F, 4 M	1 F, 5 M
	AGE	26.2 ± 7	27.7 ± 5
	HAND	All right-handed	All right-handed
MEDICAL PARTICIPANTS			
	SEX	1 F, 5 M	3 F, 3 M
	AGE	28.5 ± 3.9	28.2 ± 3.3
	HAND	5 right-handed	All right-handed
SURGICAL EXPERIENCE* (hours)	OPEN	92.7 ± 58	120.8 ± 67
	LAPARO	73.3 ± 37	90.4 ± 43
	ROBOT	0	3.3 ± 3

***Note** *The data related to the medical participants' previous experience in surgery (open surgery and laparoscopic surgery) refer both to firsthand experience and to assistance during the procedure.*

As far as it concerns the complex tasks, an evaluation of each involved ith elementary skill was calculated analogously:

$$S_i = \frac{\sum \alpha_j m_j}{\sum \alpha_j} \in [0,1] \quad (3).$$

where α_j is the weight associated to the jth metric m_j in the complex task and j ranges from 1 to 5.
The overall performance in the complex tasks was derived as the average of the skill-related ones. Table I reports the metric weights in the complex (red) and elementary (blue) tasks. These weights were assigned under expert surgeon's supervision.

D. Acquisition Protocol

A user study was carried out to test the research hypothesis. The study population consisted in 12 non-medical participants and 12 medical residents, whose specifics are reported in Table

II. All the subjects had none to little experience with robotic teleoperation. None of the volunteers had neurological disorders or visuo-motor problems with a possible negative impact on their performance. The research outline was explained to all the subjects prior to enrollment and the experiments were carried out following the recommendations of our institution. All the subjects gave written informed consent in accordance with the declaration of Helsinki.
Both the medical and non-medical users underwent the same acquisition protocol. They were randomly divided into two groups: a control group (N=6), performing the self-managed training, and an experimental group (N=6), undergoing the adaptive training. Participants were assigned by an investigator who was not involved in their recruitment; random assignment (numbers randomly generated from a computer) was blocked (block size equal to 2) and stratified by whether the participant had any medical experience. The group allocation followed the two samples block random assignment paradigm. Firstly, the subjects were introduced to the dVRK console and they were shown videos of a successful execution of each task. The total time was selected as a training constraint to have comparable protocols between the two groups. In fact, the users underwent 45 minutes of training: the ones belonging to the *control group* directly chose their exercises and they were provided with their percentage performance after each task repetition; the *experimental group* was provided with the task to perform according to the adaptive algorithm (described in the following section). Moreover, each user performed an initial baseline assessment (2 repetitions) and a final evaluation (4 repetitions) on the steady hand complex task. The former aimed at testing that all the trainees initially belonged to the same statistical population in terms of performances; the latter allowed to quantify the training effects. Finally, all the subjects went through a final test consisting of 4 repetitions of the skill transfer complex task. Before the experiments started, the users were informed about the training protocol and the modalities of final performance assessment.

E. Adaptive Logic

The adaptive logic automatically modulated the training curriculum of the experimental group as a function of their performances. The training session of these users was split into 3 units, as shown in Figure 3a. These were defined as 15 minutes training slots, where the user performed one repetition of the steady hand complex task at the beginning and then elementary tasks till the end of that unit. Once the elementary training started, the adaptive algorithm had to output the elementary task that the user should perform. The proposed method aims at maximizing the learning by filling the gaps in the different surgical skills. This was achieved by scheduling the training on the basis of a *priority index* (φ_i, where *i* corresponds to the i^{th} elementary skill) that took into account the initial user's performance in the complex task, together with the influence of the on-going elementary training. In a formal way,

$$\bar{\imath} = elementary\ task\ to\ perform = argmax(\varphi_i).$$

To initialize this index, the skill-related performance (3) in the first complex task repetition was exploited:

$$\varphi_i = 1 - S_i \quad \in [0,1] \qquad (4).$$

Once the first selected elementary task was performed, in order to update the priority index, the result in the elementary training was considered as well as the correlation of the performed elementary task with other skills and with the complex task itself. To achieve this goal, we introduced the concepts of *skill connection* and *structural significance* respectively (Figure 3b). These were quantified by computing the correlation of metric weights of the two tasks under analysis. The *skill connection* was defined as:

$$\rho_{ij} = \frac{\sum_m \beta_m^i \beta_m^j}{\sqrt{\sum_m {\beta_m^i}^2 \sum_m {\beta_m^j}^2}} \in [0,1] \qquad (5)$$

where β_m^i is the weight associated to the m^{th} metric in the i^{th} elementary task, while β_m^j is the weight associated to the m^{th} metric in the j^{th} elementary task. As far as it concerned the *structural significance*,

$$\sigma_i = \frac{\sum_m \alpha_m^i \beta_m^i}{\sqrt{\sum_m {\alpha_m^i}^2 \sum_m {\beta_m^i}^2}} \in [0,1] \qquad (6)$$

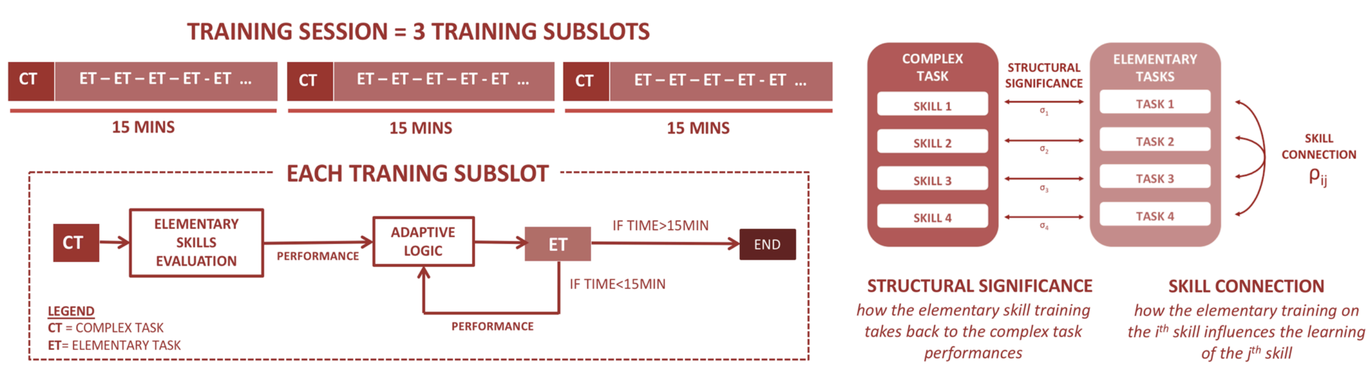

Figure 3 *a) Experimental Group Training Session. It is composed by 3 15-minutes sub-slots, each one starting with one repetition of the complex task followed by elementary tasks according to the adaptive training logic (automatic scheduling). b) The concepts of skill connection and structural significance.*

where α_m^i is the weight associated to the m^{th} metric in the complex task to evaluate the ith skill and β_m^i is the weight associated to the mth metric in the ith elementary task. In the end, the priority index was updated as

$$\varphi_i^{new} = \varphi_i^{old}\left(1 - \sigma_i\, \rho_{ij}\, P_j\right) \in [0,1] \qquad (7)$$

where j is the index of the elementary task that the user performed last time and P_j is the associated performance.

F. Statistical Analysis

Due to the small sample size, non-parametric statistical significance tests were used to compare the training effects on the two groups. The Wilcoxon rank sum test was employed to test significant differences in initial and final median performances. The performance measures were selected as dependent variables and training groups as independent factors. Statistically significant effects were assessed at $p < 0.05$. The statistical analysis was performed in MATLAB using the command *ranksum()*.

III. Results and Discussion

In this section, the main outcomes of the user study are presented and discussed, focusing on the performance comparison between the control and the experimental group. Medical and non-medical participants are compared as well. No subject gave up during the experiment and all the enrolled subjects underwent the experiment, having the related acquired data reported in this paper.

The primary outcomes of the study are depicted in Figure 4. This shows a comparison of the initial and final performances in the steady hand complex task. All the groups were characterized by an increase in the median performances and a reduction of their variability after training. The initial baseline assessment proved that all the subjects belonged to the same population (no significant difference across all the groups, see Table III). The medical users were characterized by slightly higher starting points in terms of performances but not in a significant way (similarly to the results in [47], where users with and without medical background performing with the da Vinci Surgical Simulator were compared). This can be explained by the fact that all these subjects had no prior experience with robotic surgical devices. Since the skills involved in the complex task are fundamental skills of robotic surgery, their earlier surgical experience (open and laparoscopic surgery) did not carry weight in their initial performances. This could also suggest and remark the importance of an appropriate and focused training for all the surgeons who are approaching a robotic teleoperative device.

Moving to the final evaluation of the participants in the complex task, statistically significant difference between the control and experimental group was found both in the medical

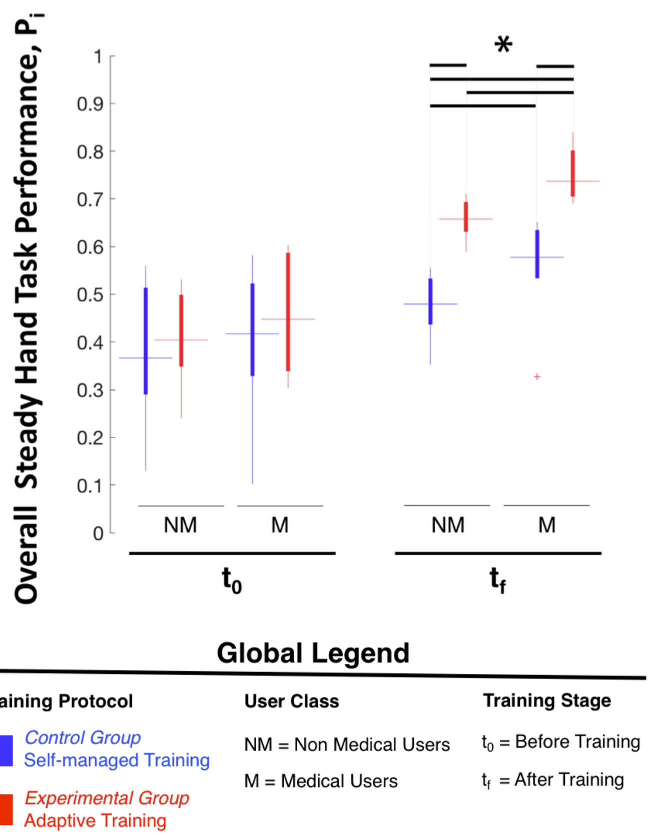

Figure 4 *Overall performances in the steady hand complex task before (t_0) and after training (t_f). The horizontal line represents the median across the population, the bottom and top edges of the bold vertical bars stand for the 25th and 75th percentiles, the light vertical bar is for whiskers and outliers are depicted as red crosses.*

Table III *Statistics about overall performances*

	t_0						t_f					
	C			E			C			E		
	m	p_{25}	p_{75}	m	p_{25}	p_{75}	m	p_{25}	p_{75}	m	p_{25}	p_{75}
NM	0.3661	0.2899	0.5138	0.4040	0.3486	0.4988	0.4793	0.4365	0.5335	0.6576	0.6308	0.6933
M	0.4171	0.3285	0.5226	0.4475	0.3389	0.5868	0.5776	0.5329	0.6344	0.7369	0.7047	0.8017

Overall Performance Median and Ranges

		NM		M	
		C	E	C	E
NM	C		**0.0022**	0.1320	**0.0022**
NM	E	0.9372		0.0649	**0.0152**
M	C	0.6991	1		**0.022**
M	E	0.2403	0.5887	0.5887	

(t_0 light grey, t_f dark grey)

Statistical tests

Note *On the left, the table reports the medians (m) and percentiles (p_{25}, p_{75}) associated to the overall performances in the steady hand task. The initial performances(t_0) are shown in light grey, while the final performances (t_f) are in dark grey. On the right, each box of the table contains the p-value of the Wilcoxon rank sum test between the performances of the populations of the corresponding rows and columns. Statistically significant effects (bold) were assessed for p<0.05.*

Figure 5 *A) Training curriculum average composition (i.e. number of tasks performed for each category); B) Percentage of task switching (i.e. number of times the user performed consecutively two different tasks normalized by the total number of performed tasks). Same legend and boxplot details with respect to Fig.4 apply.*

and non-medical class: the subjects who underwent the adaptive training achieved higher final performances compared to the ones who self-managed their training.

Comparing the non-medical and medical participants, similar results were identified: the adaptive training enabled higher final performances in both the classes. Going further in detail, the medical subjects were characterized by higher learning effects (estimated as the median performance variations, $p_{final} - p_{initial}$) when compared to the non-medical users: 11.31% (NM) vs. 16.05% (M) in the control groups and 25.36% (NM) vs. 28.94% (M) in the experimental groups. This could suggest a better predisposition to learning the robotic surgical skills by the subjects with a medical background, even if their baseline performances were comparable to the ones of the non-medical participants. Additionally, the motivational factor could play a key role in learning, thus boosting the improvement of subjects from the medical sphere.

The analysis of Figure 4 leads to the key consideration that the adaptive training allowed both the medical and non-medical participants to achieve higher learning effects and, in turn, final performances compared to the participants who directly managed their training session. The overall composition of the training curriculum helps in understanding these primary outcomes. Figure 5A reports the average number of task repetitions across the users of each class (medical or not) and group (control and experimental). The total number of task repetitions was higher in the experimental group, both for medical and non-medical participants. The immediate explanation of this lies in the fact that the experimental subjects had just to perform the task the adaptive algorithm selected for them; on the other hand, the users belonging to the control group had to think about which task executing next on the basis of their experience and the performance feedback they were provided with. This decisional step implies a loss of time and, in turn, a lower number of task performed in a given time window. Considering the time constraints related to the usage of simulators by medical residents, these results highlight the promising feature of an automatic schedule of the training in order to take full advantage of the available time with the device.

Additionally, the control group performed a higher number of complex task repetitions with respect to the experimental users in the case of non-medical participants, while an opposite trend was identified among the medical subjects. Considering that the average duration of the complex task (93 sec, across all the novice users) was higher than the average duration of the elementary tasks (75 sec, across all the novice users), the abovementioned trends could explain why the difference in total number of tasks between the control and the experimental subjects was less marked among the medical users.

Figure 5B shows the *percentage of task switching*, that is defined as the number of times the user performed consecutively two different tasks normalized by the total number of performed tasks. Statistically significant difference was found between the control and experimental users of each class: while the participants who self-managed their training moved towards focusing on a single exercise till proficiency before switching, the adaptive logic pushed towards frequent switching and consecutive tasks diversification.

In synthesis, a preliminary explanation of the higher end points that the experimental users achieved could be the optimization of the training time. At the same time, the higher final performances of the experimental users can be linked to the reduction of the subject's cognitive load during training, when the subject could focus just on the exercise and not also on the choice of the task. However, a more detailed analysis of the effects the adaptive training managed to obtain is required: Figure 6 shows the initial and final skill-related performances of the different groups. Again, these are the performances associated to each one of the involved skills (3D Perception, Hand-Eye Coordination, Wrist Articulation and Object Exchange) achieved during the baseline assessment and the final test on the steady hand task. Analogously to Figure 4, the general trend is an increase in the median performances and a reduction of their variability from the initial to the final experimental time step. As reported in Table IV, the comparison of skill performances between control and experimental group puts into evidence the absence of significant difference pre-training and the achievement of higher performances in each skill by the participants who underwent the adaptive training. Additionally, the analysis of the distribution of the four skills across a single group highlights as the initial dispersion of all the performances, was better compensated by the experimental users. This can be

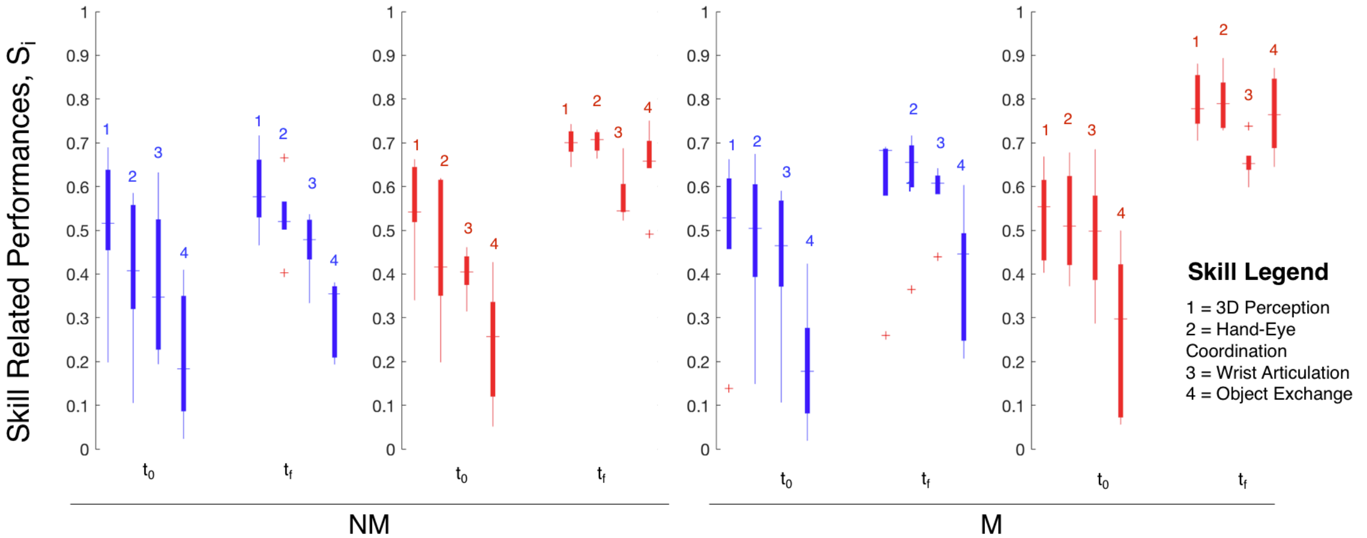

Figure 6 *Initial and final performances S_i in the steady hand complex task related to the single involved skills (3D perception, hand-eye coordination, wrist articulation and object exchange). Same legend and boxplot details with respect to Fig.4 apply.*

quantified by introducing an *inter-skill variance* (defined as the difference between the 75th and 25th percentiles of the population composed by all the final skill performances of a certain group). This parameter was lower in the experimental group (non-medical: 0.09 vs. 0.17; medical: 0.15 vs. 0.21) after training. Focusing again on the skill spread before training, the results are similar across all the groups (lowest performances in skill 4, i.e. the object exchange).
Moving to the end of the training session, that biggest skill gap was compensated just in the experimental users. These results can be interpreted by assuming that the adaptive training achieved its objective of identifying the subject's gaps in each skill and structuring the curriculum to fill those gaps.

A final secondary outcome is derived from the examination of the performances achieved in the *skill transfer task* at the end of the experiment (Figure 7). This exercise involves the same skills of the steady hand task and it was totally new to the users when they were asked to perform it. The statistical analysis of

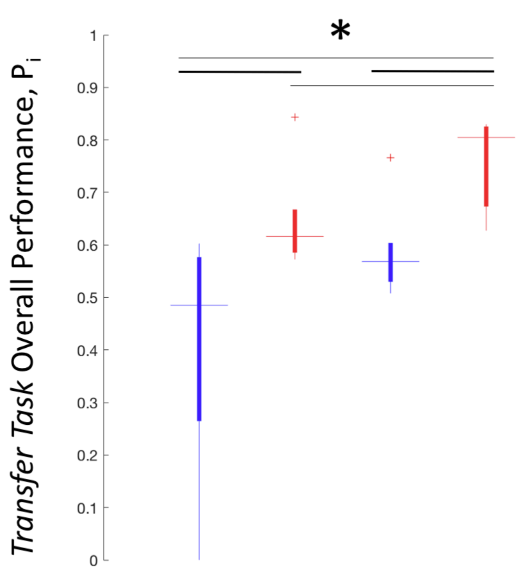

Figure 7 *Post-training overall performances in the skill transfer task. Same legend and boxplot details with respect to Fig.4 apply.*

Table IV

SKILL COMPARISON: CONTROL vs ADAPTIVE				
	BEFORE TRAINING		AFTER TRAINING	
SKILL	NM	M	NM	M
1	0.7488	0.8728	**0.0374**	**0.0039**
2	0.6310	0.7488	**0.0065**	**0.0039**
3	0.5218	0.6310	**0.0104**	**0.0250**
4	0.6310	0.6310	**0.0039**	**0.0039**

Note *Each box contains the p-value of the Wilcoxon rank sum test between the performances of control and experimental groups in each skill.*

the performances of the different groups highlights that the experimental participants performed better with respect to the control users (both in the non-medical – $p=0.0286$ – and medical – $p=0.0152$ – classes). Again, this could be explained by assessing that the skill-oriented and adaptive algorithm managed to train the users on the skills, not just on a certain task that involves them, as well as it was able to evaluate the subjective gaps and propose a user-specific training schedule in order to maximize the learning benefit in a fixed training time with the simulator.

IV. Conclusion

This study introduced a skill-oriented and adaptive curriculum for training in robotic surgery using virtual reality simulators. This novel approach was compared to a traditional self-managed training, where the trainee decides which tasks to perform on the basis of his/her experience and a performance feedback about his/her previous attempts. A feasibility study was carried out on a population without medical background and on a sample of medical residents.

To sum up the results, the primary outcome of the study was the verification of the research hypothesis: subjects undergoing adaptive training achieved better final performances compared to the control group (i.e. subjects who self-managed their training session).

Additionally, this work led to the following secondary considerations:

- the adaptive training was capable to better identify and compensate for the trainee's gaps in the involved skills;
- the subjects who underwent the adaptive training managed to optimize the available time with the simulator by performing a higher number of tasks;
- skill transfer to a new simulated task applied most to experimental users, showing that the adaptive training strategy managed to better teach the skills instead of the task itself. This is the real goal of a simulation-based training program (i.e. acquire the fundamental skills in robotic surgery);
- all the above-mentioned considerations are valid for both the non-medical and medical population;
- no significant difference between the two classes was derived from the baseline assessment. This could be due to the fact that medical residents had no prior experience with robotic surgery. This kind of surgery implies new control modalities with respect to conventional surgical approaches and dealing with these robotic devices is not immediate in terms of motor learning. These results are in line with the outcomes that other investigations have previously derived regarding skill transferability among heterogeneous surgical approaches [48]. Following this reasoning, an appropriate training constitutes a fundamental step towards a safe introduction of these technologies in the surgical scenario.

In synthesis, the proposed adaptive approach led to promising results, showing how a *smart* training schedule can maximize the learning process when training with a simulator. However, this work just focused on basic skills in human robot interaction. These skills are fundamental for trainees who first approach surgical robotics but how they impact on clinical outcomes was not under investigation in this work.

In response to the lack of supervised curricula for VR simulators, the automated and objective method that we designed could suggest the introduction of an artificial intelligence mentor as an integrated component of a simulation-based platform for training in robotic surgery. This way, each training session can be guided by a constant and objective performance evaluation, followed by an *ad hoc* scheduling of the intra-session curriculum.

The current study was a proof of concept and several factors have still to be investigated. Among the major limitations that characterized the work, this research focused just on the initial part of the trainees' learning curve and longer or multiple session experiments should be performed. Equally, this research considered a limited set of fundamental skills of robotic surgery: a detailed identification of these basic capabilities and an analysis of a broader skills spectrum have to be carried out. This should be combined with an upper level recognition of the surgical tasks that involve these skills and with a clear understanding about how these skills map onto the complex tasks and clinical outcomes. Additionally, the skill assessment method could be refined in order to generalize the metric selection and the weight computation towards performance assessment. In particular, the definition of weights should be achieved by using systematic methods like the Delphi process to take into account experts' opinion and to reach experts' consensus (as in [49]). Extensive analysis of expert data should be included in order to provide solid validity of the simulation-based training protocol.

Finally, a wider population has to be analyzed to move from a feasibility study towards a validation work. Larger sample size could lead to higher statistical determinism, as well as more restrictive selection criteria (or partition according to single surgical specialties) of the medical population could result in interesting considerations.

In terms of additional future developments, the integration of methods capable to quantify the trainee's performance in a task-independent way could be an added value to this research. Machine learning-based approaches, which can exploit the data (both associated to the master kinematics, as in several previous studies [50-52], and to the interaction with the virtual environment) recorded during these preliminary experiments in order to assess the user's proficiency, are currently under our investigation.

In the direction of an adaptive training protocol that can guide the trainee through the learning of a full robotic surgical procedure, further analyses could involve a control group that performs the training under the guidance of a real mentor. This experimental protocol will assess the efficacy of the adaptive algorithm with respect to an expert but subjective proctoring (the current golden standard when learning surgical procedures).

ACKNOWLEDGMENT

We would like to thank all the participants to the experimental sessions. In particular, we are grateful to the medical residents who kindly took part to the study from Policlinico di Milano, Ospedale San Giuseppe di Milano and Istituto Europeo di Oncologia.

REFERENCES

[1] A. Aydin, K. Ahmed, A. M. A. Shafi, M. S. Khan, P. Dasgupta, "The role of simulation in urological training - A quantitative study of practice and opinions", *Surgeon*, 14: 301–7, 2016.

[2] D. M. Gaba, "The future vision of simulation in health care", *Qual Saf Health Care*, 13(Suppl. 1): i2–i10, 2004.

[3] A. Ziv, P. R. Wolpe, S. D. Small, S. Glick, "Simulation-based medical education: an ethical imperative", *Acad Med: J Assoc Amer Med Coll*, 78(8):783–788, 2003.

[4] V. N. Palter, N. Orzech, R. Aggarwal, A. Okrainec, T. P. Grantcharov, "Resident perceptions of advanced laparoscopic skills training", *Surg Endosc*, 25(7):2408–2410, 2011.

[5] S. Maeso, M. Reza, J. A. Mayol et al, *Efficacy of the Da Vinci Surgical System in abdominal surgery compared with that of laparoscopy: a systematic review and meta-analysis*, *Ann Surg.* 252: 254-262, 2010.

[6] W. R. J. Chitwood, L. W. Nifong, W. H. H Chapman et al, "Robotic surgical training in an academic institution", Ann Surg. 234: 475–486, 2001.

[7] J. Rassweiler, M. Hruza, D. Teber, L.M. Su, "Laparoscopic and robotic assisted radical prostatectomy - critical analysis of the results", *Eur Urol*. 49:612-624, 2006.

[8] A. R. Lanfranco, A. E. Castellanos, J. P. Desai et al., "Robotic surgery: a current perspective", *Ann Surg*. 239:14–21, 2004.

[9] J. E. Thompson, S. Egger, M. Bohm, et al. "Superior quality of life and improved surgical margins are achievable with robotic radical prostatectomy after a long learning curve: a prospective single- surgeon study of 1552 consecutive cases", *Eur Urol*, 65:521–31, 2014.

[10] I. Nisky, M. H. Hsieh, and A. M. Okamura, "Uncontrolled manifold analysis of arm joint angle variability during robotic teleoperation and freehand movement of surgeons and novices", *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 12, pp. 2869-2881, 2014.

[11] A. M. Okamura, "Haptic feedback in robot-assisted minimally invasive surgery", *Current Opinion in Urology,* vol. 19, no. 1, pp. 102-107, 2009.

[12] B. S. Farivar, S. Behzad, M. Flannagan and I. M. Leitman. "General Surgery Residents' Perception of Robot-Assisted Procedures During Surgical Training," *Journal of Surgical Education*; 72(2): 235–42, 2015.

[13] R. A. Fisher, P. Dasgupta, A. Mottrie, et al. "An over-view of robot assisted surgery curricula and the status of their validation", *Int. J. Surg*.; 13:115–23, 2015.

[14] T. J. Guzzo, M. L. Gonzalgo, "Robotic surgical training of the urologic oncologist*", Urol Oncol*, Mar-Apr;27(2):214-7, 2009.

[15] A. Moglia, V. Ferrari, L. Morelli, M. Ferrari, F. Mosca, & A. Cuschieri, "A systematic review of virtual reality simulators for robot-assisted surgery", *European urology*, 69(6), 1065-1080, 2016.

[16] J. D. Bric, D. C. Lumbard, M. J. Frelich, & J. C. Gould, "Current state of virtual reality simulation in robotic surgery training: a review". *Surgical endoscopy*, *30*(6), 2169-2178, 2016.

[17] F. J. Carter, M. P. Schijven, et al., "Consensus guidelines for validation of virtual reality surgical simulators", *Surg Endosc*., Dec;19(12):1523-32, 2005.

[18] A. J. Hung, P. Zehnder, M. B. Patil, et al., "Face, content and construct validity of a novel robotic surgery simulator". *J Urol*, 186(3): 1019–1024, 2011.

[19] M. A. Lerner, M. Ayalew, W. J. Peine, C. P. Sundaram, "Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system?", *J Endourol*, 24(3):467–472, 2010.

[20] S. Messick, "Validity", in R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-104), 1989.

[21] N. J. Borgersen, T. M. Naur, S. M. Sørensen, F. Bjerrum, L. Konge, Y. Subhi, & A. S. S. Thomsen*, "Gathering validity evidence for surgical simulation: a systematic review",* Annals of surgery, 267(6), 1063-1068, 2018.

[22] J. C. Jr. Rosser, L. E. Rosser, S. S. Raghu, "Objective evaluation of a laparoscopic surgical skill program for residents and senior surgeons", *Arch Surg.* 133: 657–661, 1998.

[23] K. Moorthy, Y. Munz, S. K. Sarker, A. Darzi, "Objective assessment of technical skills in surgery", *BMJ*; 327:1032–7, 2003.

[24] S. D. Herrelland, J. A. Smith, "Robotic-assisted laparoscopic prostatectomy: what is the learning curve?", *Urology*, 66 (5 Suppl):105-7, 2005.

[25] M. Connolly, J. Seligman, A. Kastenmeier, M. Goldblatt, & J. C. Gould, "Validation of a virtual reality-based robotic surgical skills curriculum", *Surgical endoscopy*, 28(5), 1691-1694, 2014.

[26] T. P. Grantcharov, V.B. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, J., & P. Funch-Jensen, "Randomized clinical trial of virtual reality simulation for laparoscopic skills training". *British journal of surgery*, *91*(2), 146-150, 2004.

[27] T. S. Lendvay, P. Casale, R. Sweet, & C. Peters, "Initial validation of a virtual-reality robotic simulator". *Journal of Robotic Surgery*, *2*(3), 145-149, 2008.

[28] Moorthy K, Munz Y, et al., "Dexterity enhancement with robotic surgery"*, Surg Endosc.*, 18(5):790-5, 2004.

[29] N. Enayati, G. Ferrigno, and E. De Momi, "Skill-based human–robot cooperation in tele-operated path tracking", *Autonomous Robots*, 2017.

[30] J. MacDonald, R. G. Williams, D. A. Rogers, "Self-assessment in simulation-based surgical skills training", *The American Journal of* Surgery, Volume 185, Issue 4, 319 – 322, 2003.

[31] D. A. Davis, P. E. Mazmanian, M. Fordis, R. Van Harrison, K. E. Thorpe, L. Perrier, "Accuracy of Physician Self-assessment Compared With Observed Measures of Competence: A Systematic Review", *JAMA*; 296(9): 1094–1102, 2006.

[32] K. F. Kowalewski, J. D. Hendrie, M. W. Schmidt, C. R. Garrow, T. Bruckner, T. Proctor, *et al.* , "Development and validation of a sensor-and expert model-based training system for laparoscopic surgery: the iSurgeon", *Surgical endoscopy,* 31(5), 2155-2165, 2017.

[33] A. Malpani, S. S. Vedula, H. C. Lin, G. D. Hager, & R. H. Taylor, Effect of real-time virtual reality-based teaching cues on learning needle passing for robot-assisted minimally invasive surgery: a randomized controlled trial, *International journal of computer assisted radiology and surgery*, 2020.

[34] G. Rauter , N. Gerig, R. Sigrist, R. Riener, P. Wolf, "When a robot teaches humans: Automated feedback selection accelerates motor learning", *Science Robotics* 4.27: eaav1560, 2019.

[35] N. Enayati, A. M. Okamura, A. Mariani, E. Pellegrini, M. Coad, G. Ferrigno, and E. De Momi, "Robotic Assistance-as-Needed for Enhanced Visuomotor Learning in Surgical Robotics Training: An Experimental Study", *International Conference on Robotics and Automation (ICRA)* 2018.

[36] C. Shih, H. Y. D. Lu, L. Sun, J.-T. Huang, J. Packard, "An adaptive training program for tone acquisition", *Speech Prosody*, 100981, Chicago, pp. 1–4, 2010.

[37] C. R. Kelley, "What is adaptive training?", *Human Factors*, 11(6), 547–556., 1969.

[38] M. A. Guadagnoli, D. L. Timothy, "Challenge Point: A Framework for Conceptualizing the Effects of Various Practice Conditions in Motor Learning", *Journal of Motor Behavior*; 36(2): 212–24, 2004.

[39] A. Mariani, E. Pellegrini, N. Enayati, P. Kazanzides, E. De Momi, "Design and Evaluation of a Performance-based Adaptive Curriculum for Robotic Surgical Training: a Pilot Study", *Annual International Conference of the IEEE Engineering in Medicine and Biology Society Conference 2018*: 2162-2165, 2018.

[40] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. Dimaio, "An open-source research kit for the da Vinci Surgical System," Proc. - *IEEE Int. Conf. Robot. Autom*., pp. 6434–6439, 2014.

[41] N. Enayati, A. Mariani, E. Pellegrini, T. Chupin, G. Ferrigno, and E. De Momi, "A Framework for Assisted Tele-operation with Augmented Reality," in *CRAS: Joint Workshop on New Technologies for Computer/Robot Assisted Surgery*, 2017.

[42] E. Coumans, "Bullet physics library," Open source bulletphysics.org, 2013.

[43] K. Mamou and F. Ghorbel, "A simple and efficient approach for 3D mesh approximate convex decomposition," in 2009 *16th IEEE International Conference on Image Processing (ICIP),* pp. 3501–3504, 2009.

[44] G. Dulan, R. V. Rege, D. C. Hogg et al., "Developing a comprehensive, proficiency based training program for robotic surgery", *Surgery*;152:477–88, 2012.

[45] A. J. Hung, M. B. Patil, P. Zehnder, J. Cai, C. K. Ng, M. Aron, ... , M. M Desai, Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study, *The Journal of urology*, 187(2), 630-637, 2012.

[46] A. Blavier, Q. Gaudissart, G. B. Cadière, A. S. Nyssen, Comparison of learning curves and skill transfer between classical and robotic laparoscopy according to the viewing conditions: implications for training, *The American journal of surgery*, 194(1), 115-121, 2007.

[47] A. Moglia, V. Ferrari, F. Melfi, M. Ferrari, F. Mosca, A. Cuschieri, L. Morelli, "Performances on simulator and da Vinci robot on subjects with and without surgical background", *Minimally Invasive Therapy and Allied Technologies,* 0(0), 1–6, 2017.

[48] K. F., Kowalewski, Karl-Friedrich, et al. "Skills in minimally invasive and open surgery show limited transferability to robotic surgery: results from a prospective study." Surgical endoscopy 32.4: 1656-1667, 2018.

[49] M. W. Schmidt, K. F. Kowalewski, M. L. Schmidt, E. Wennberg, C. R. Garrow,. S. Paik, , ... & F. Nickel, The Heidelberg VR Score: development and validation of a composite score for laparoscopic virtual reality training. *Surgical endoscopy*, 33(7), 2093-2103, 2019.

[50] R. Aggarwal, T. Grantcharov, K. Moorthy, T. Milland, P. Papasavas, A. Dosis, A. Darzi, "An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room", *Annals of Surgery*, 245(6), 992–999., 2007.

[51] Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, G. D. Hager, "JHU-ISI Gesture and Skill Assessment Working Set (JIGSAWS): A Surgical Activity Dataset for Human Motion Modeling. Modeling and Monitoring of Computer Assisted Interventions (M2CAI)" – *MICCAI Workshop*, 1–10., 2014.

[52] Z. Wang, A. Majewicz Fey, "Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery". *International Journal of Computer Assisted Radiology and Surgery,* 13(12), 1959–1970., 2018.