Effect of real-time virtual reality-based teaching cues on learning needle passing for robot-assisted minimally invasive surgery: a randomized controlled trial

Anand Malpani 1,2 · S. Swaroop Vedula 1 · Henry C. Lin 2 · Gregory D. Hager 1 · Russell H. Taylor 1

Received: 16 November 2019 / Accepted: 3 April 2020 © CARS 2020

Abstract

Purpose Current virtual reality-based (VR) simulators for robot-assisted minimally invasive surgery (RAMIS) training lack effective teaching and coaching. Our objective was to develop an automated teaching framework for VR training in RAMIS. Second, we wanted to study the effect of such real-time teaching cues on surgical technical skill acquisition. Third, we wanted to assess skill in terms of surgical technique in addition to traditional time and motion efficiency metrics.

Methods We implemented six teaching cues within a needle passing task on the da Vinci Skills Simulator platform (noncommercial research version). These teaching cues are graphical overlays designed to demonstrate ideal surgical technique, e.g., what path to follow while passing needle through tissue. We created three coaching modes: TEACH (continuous demonstration), METRICS (demonstration triggered by performance metrics), and USER (demonstration upon user request). We conducted a randomized controlled trial where the experimental group practiced using automated teaching and the control group practiced in a self-learning manner without automated teaching.

Results We analyzed data from 30 participants (14 in experimental and 16 in control group). After three practice repetitions, control group showed higher improvement in time and motion efficiency, while experimental group showed higher improvement in surgical technique compared to their baseline measurements. The experimental group showed more improvement than the control group on a surgical technique metric (at what angle is needle grasped by an instrument), and the difference between groups was statistically significant.

Conclusion In a pilot randomized controlled trial, we observed that automated teaching cues can improve the performance of surgical technique in a VR simulator for RAMIS needle passing. Our study was limited by its recruitment of nonsurgeons and evaluation of a single configuration of coaching modes.

Keywords Virtual reality · Surgical training and assessment · Surgical robots · Randomized controlled trial · Evaluation and validation

Introduction

Technical skill training for surgeons is important because poor skill is associated with more adverse outcomes for patients [2,9]. Simulation-based training has been advocated by the surgical community to address patient safety concerns that arise from learning in the operating room (OR) [11,25]. Simulation training allows for repeated and focused practice of skills. Likewise, introduction of new technology in the OR, e.g., robotic devices, requires transfer of existing skills onto the new platform and has a learning curve.

1 Henceforth, in this paper, surgical skill will refer to technical skill specifically.
As new surgical robots and advanced versions of existing ones enter the OR, there is a need for standardized training, assessment, certification, and privileging to ensure patient safety [15,19,20]. Research studies have shown the effectiveness of simulation training in skill acquisition and transfer of skills to the OR for robot-assisted minimally invasive surgery (RAMIS) using the da Vinci® Surgical System. However, the uptake of simulation training by trainee and practicing surgeons has been low [27,29]. Simulation training lacks expert teaching and coaching that is accessible in the OR since practicing surgeons cannot be present at simulation practice sessions due to time constraints. Thus, there is a need for development of automated and standardized teaching and coaching for RAMIS skills.

Teaching and coaching comprise various activities such as assessment, demonstration, critiquing, and recommending deliberate practice [1,16]. Virtual reality-based (VR) trainers have gained popularity in RAMIS [18]. They provide automated, objective, and structured assessment to the learner using summative metrics that measure time, motion efficiency, force sensitivity, and error counts. Such assessment provides information to the learner about “how did I do”. However, the learning remains incomplete if this is not coupled with feedback that answers the questions “how do I do it correctly” and “what was I doing wrong”. Feedback and teaching that focuses on identifying, explaining, and demonstrating errors has proven learning value in surgical simulation training [10]. Thus, we choose to focus on teaching surgical technique for fine-grained actions termed critical learning elements (CLE). These learning elements are critical in reducing errors and improving task proficiency.

Since VR simulations are computer-generated, they provide a structured environment to enable intelligent systems to provide automated teaching and coaching that can augment learning in the absence of surgeon educators [21]. In this paper, we present our approach using VR training toward enabling two components of such automated teaching, i.e., demonstration of surgical technique and assessment of surgical technique. Our contributions in this paper are: (1) real-time teaching cues for learning surgical technique of needle passing in a RAMIS setting, (2) metrics that measure surgical technique, and (3) a pilot evaluation of the framework using the da Vinci® Skills Simulator™ (dVSim; Intuitive Surgical Inc., USA) in a randomized controlled trial.

### Automated teaching framework

Our software framework comprises teaching cues, coaching modes, a task progress manager, and a module for data logging and computing performance metrics. While this framework can apply to various RAMIS skills, we will be focusing on the skill of needle passing (NP) in this manuscript.

### Needle passing task

We modified an existing NP task (named “Posterior Around the World” commercially) on a research version of the dVSim, a VR training platform for RAMIS. In this task (Fig. 1a, Electronic Supplementary Material 1), the user is provided two “Large Needle Driver” instruments to pass a curved needle across a deformable tissue at eight locations arranged around a circle in a clockwise sequence. Each location is pre-marked by a pair of circular NP targets along the radius such that the entry target is located on the inner side and the exit target is on the outer side. The needle is to be passed into the tissue through the entry target and exited through the exit target. The current targets are highlighted in yellow with the entry highlighted with a flashing yellow. Else, the task does not proceed to highlight the next pair of targets.

### Teaching cues

Our framework introduces teaching cues as teaching and feedback tools that demonstrate ideal behavior at CLE (critical learning elements) of a surgical skill (Fig. 1). In the case of NP, we focus on CLE aimed at the surgical technique of loading (grasping) the needle, inserting needle into a tissue, and passing it through the tissue, based on guidelines from the ACS/APDS Surgery Resident Skills Curriculum. We referred to the Phase I modules 3, 13, and 14 that focus on suturing skills. A video showing the teaching cues described below can be found as Electronic Supplementary Material 1.

### Ideal instrument indicator

**Purpose** This indicates the instrument (left v/s right) that would be ideal for inserting needle to perform current NP segment.

**Significance** Using a nonideal instrument to insert needle into the tissue may lead to constrained ergonomics, awkward needle insertion angles, and excessive stress on tissue.

**Appearance** This is indicated using red-colored spheres that are positioned at the instrument jaw tips (Fig. 1a).

**Computation** This can be determined using the instrument’s initial setup pose, current NP targets pose, and joint limits on the robot manipulator holding the instrument. In the case of NP instances where either instrument would be suited, the user’s handedness is preferred.

---

Scott et al. [22].
Grasp position guide

**Purpose** This indicates the region along the needle’s curved body that is ideal for grasping to perform needle insertion for current NP segment.

**Significance** Grasping too close to the tip may require re-grasping to continue driving needle to reach the exit target location. Grasping too far may result in poor control of needle and excessive force on tissue.

**Appearance** Two flashing yellow spherical overlays along the needle’s curvature indicate the grasping region (Fig. 1a).

**Computation** In the current framework, we selected a fixed value for the bounds, viz. $135^\circ$ and $165^\circ$ respectively$^2$, where the tip of the needle is located at $0^\circ$. Ideally, this range should be computed using distance between entry and exit targets, and needle curvature radius.

Grasp orientation guide

**Purpose** This indicates the ideal orientation of the instrument’s jaws with respect to the needle curvature for the current NP segment.

**Significance** Deviating from the ideal may result in excessive lateral force exerted by the needle body on the tissue at the insertion point during NP, since articulation of the instrument wrist gets constrained.

**Appearance** A light-green semitransparent copy of instrument’s jaws attached to the needle body indicates the ideal orientation (Fig. 1a). The guide’s transparency increases as instrument approaches ideal pose, eventually disappearing.

**Computation** The framework positions this at the center of the ideal grasping region ($150^\circ$) and at right angles to the needle plane$^2$. The pointing direction of jaws is determined using the current NP direction and the suggested ideal instrument by the first cue.

Ideal drive path overlay

**Purpose** This displays the ideal path that the needle’s tip and body should trace through during current NP segment.

**Significance** Driving the needle along its curvature results in minimal lateral forces at insertion and exit locations as well as within the tissue.

**Appearance** A cyan arc passing through current NP targets and pointing in the drive direction (Fig. 1a) shows the ideal path.

**Computation** The ideal path has the same radius of curvature as the needle. The path’s center of curvature is positioned such that the arc passes through center of current NP targets.

Trajectory playback overlay

**Purpose** This replays the needle tip’s trajectory from the NP attempt that gets completed.

**Significance** Showing the NP attempt immediately post-completion may help self-reflect and provide visual feedback on deviation from ideal path.

**Appearance** A red curve shows needle tip trajectory along with the ideal path as a cyan arc (Fig. 1b). A green sphere
traces the motion at real-time speed. This is projected above the tissue surface for better viewing.

Computation The framework logs the needle tip trajectory during a NP attempt along with time stamps. The orientation of cue is determined using the view direction of the endoscope.

Video demonstration overlay

Purpose This shows a video playback of a previously recorded expert/ideal performance of the current NP segment.

Significance Demonstrating ideal performance of current NP can enable focused real-time learning and show ideal behavior at CLE.

Appearance A static rectangle textured with the demonstration movie on an infinite playback loop is used. The overlay can be toggled to appear on the focal plane (Fig. 1d) or at the in situ location (Fig. 1c). This is done by bringing an instrument tip close to the (video) icon.

Computation The appropriate NP segment video recording is shown based on current state of the task progress manager.

Coaching modes

Since learners have varying levels of expertise, we implemented three coaching modes that make our framework adapt to the learner’s expertise.

1. TEACH—a complete hands-on teacher
   The coach begins by demonstrating the different steps of the task and presenting tips on them that lead to expert performance. Each of the teaching cues appears as the task progresses.

2. METRICS—a mentor that intervenes if needed
   The coach monitors performance and intervenes only when performance falls below certain thresholds on metrics such as instrument path length, time elapsed, excessive force applied on tissue, and excessive needle tissue pierces. Relevant teaching cues and text prompts explaining the reason for intervention appear for that particular NP task segment only.

3. USER—a hands-off guide
   The coach provides guidance and mentoring only when the trainee requests for it using a (bulb) icon as shown in Fig. 1d. Teaching cues specific to the current NP task segment appear.

Task progress manager

This component of our framework enables and disables the various teaching cues by reading the current state of the VR simulation and the current coaching mode. The task flow can be depicted using a directed graph, as shown in Fig. 2 we revised Fig. 2. for the needle passing task described in “Needle passing task” section. The nodes represent task state, i.e., needle insertion, needle extraction, etc, while the edges represent a sequence of interactions that result in state transitions. For example, in a needle passing task, interactions occur between needle driver and needle (e.g., grasping), and between needle and tissue (e.g., insertion). This sequence of interactions results in a transition in the task flow from State 0 (where the needle was completely outside the tissue) to State 1 (where the needle was inserted into the correct entry target). The task progress manager relies on the current task state, interactions between objects, and the task-specific programmed logic to monitor task progress.

The initial four cues (Fig. 1a) appear in the scene before the user starts driving the needle through the tissue (during State 0). These cues guide the user to set up their instruments and needle in the ideal position. Once the needle pierces the tissue (start of State 1), the ideal tool, grasp position, and grasp orientation cues disappear. The ideal drive path overlay remains in view to guide needle driving, and it disappears once the user pulls the needle out of the exit target (end of State 3). Following that, the trajectory playback overlay appears with a preset timer (10 seconds). The user is provided a ‘dismiss’ icon (red cross) to make the playback cue disappear sooner (Fig. 1b). The video demonstration cue is visible throughout the task execution at the in situ location (Fig. 1c). A video showing these teaching cues in action is available as Electronic Supplementary Material 1.

Metrics on surgical technique

The teaching cues are targeted at demonstrating ideal surgical technique. For assessment of surgical technique, we introduce deficit metrics that measure deviations from such ideal behavior. These metrics are used to trigger teaching cues in the METRICS mode as well.

- Grasp Position Deviation: average deviation in grasp location from the ideal location recommended by the Grasp Position Guide cue (150°),
- Grasp Orientation Deviation: average deviation in grasps direction from right angles to the needle plane (Grasp orientation guide cue),

3 Illustrations are provided as Electronic Supplementary Material 2, 3, and 4.
Three coaching modes in our framework—teaching, independent, self-driven practice (control group). All participants were assigned randomly to either learning from our framework (experimental group) or through independent, self-driven practice (control group). We used computer-generated random numbers to assign study participants, stratified by whether they were engineers or clinical trainers, to either learning from our framework (experimental group) or through independent, self-driven practice (control group). We used computer-generated random numbers to assign study participants, stratified by whether they were engineers or clinical trainers, to either learning from our framework (experimental group) or through independent, self-driven practice (control group).

We conducted a randomized controlled trial to study the effect of our automated teaching framework on skill acquisition for NP in RAMIS. We obtained ethical approval for our study (Western IRB; Protocol 20121049) and conducted it at Intuitive Surgical Inc. (Sunnyvale, California, USA).

We recruited study participants from two cohorts—clinical trainers and engineers, who were employed at Intuitive Surgical, Inc. We used computer-generated random numbers to assign study participants, stratified by whether they were engineers or clinical trainers, to either learning with our framework (experimental group) or through independent, self-driven practice (control group). All participants filled a pre-study questionnaire about previous experience with the da Vinci® systems. Then, they performed a baseline trial of the NP task. Subsequently, participants in the experimental group practiced the task once, under each of the three coaching modes in our framework—TEACH, METRICS, and USER (in this specific order). Participants in the control group independently practiced the task three times without automated teaching. Immediately after this, all participants performed a final test trial without automated teaching. Following this, they responded to a post-study questionnaire on the clarity and quality of feedback, perceived effectiveness of our teaching framework, and a self-assessment of their performance.

We computed performance metrics ($M$) focused on time, motion efficiency, errors, and surgical technique ("Metrics on surgical technique" section). Motion efficiency features were based on [8,14]. Errors included the number of needle pierces in tissue, excess force exerted by instruments on other objects, and excess force exerted by needle on the tissue. An empirical force threshold was set to measure duration and count of excessive force application by instruments and needle. We imputed missing data using the mean for continuous variables and median for count variables from the data on participants within the same group and the same trial (baseline, final, or respective practices). We compared the change in the value of a metric from baseline ($\Delta M = M_{\text{current}} - M_{\text{baseline}}$) between the groups using a Mann–Whitney $U$ test. We performed this test for the final repetition as well as for each of the three in-between practice repetitions. For illustration purposes, effect size values for $\Delta M$ were calculated for each metric as per the Cohen’s $d$ statistic [5].

Results

We assigned 16 participants to each group; two participants in the experimental group did not complete the study and were dropped out. Six participants had incomplete data due to technical reasons (experimental: 2 and control: 4). We analyzed data from 30 participants (experimental: 14, control: 16) of which six were clinical trainers (experimental: 3, control: 3). Control and experimental groups had similar previous experience using the da Vinci® systems. In the post-study questionnaire, most participants (93.3%) perceived an improvement in their final performance relative to the baseline. The experimental group (≥ 85%) rated all but one of the teaching cues as “intuitive,” “clear to understand,” and “effective for learning.” They found the trajectory playback cue to be not intuitive, not easy to understand, and not effective for learning (22% negative rating, 22% neutral rating). Ninety-two percent of the experimental group felt that such feedback is essential for effective learning in both the presence and absence of a surgical educator or mentor. The control group (68%) felt that real-time teaching and feedback would have helped them in improving their performance. While the control group was equivocal about the effectiveness of real-time feedback (in general), 93% of them preferred to have such feedback for themselves.
We observed statistically significant difference in the improvement in performance ($\Delta M$) between experimental and control groups on one metric—Grasp Orientation Deviation (Table 1; smaller values indicate greater learning). Figure 3 shows the difference (effect size values) in performance improvement over the baseline between experimental and control groups. Time and motion efficiency metrics uniformly show higher learning in control group (warm colors), while deficit metrics show higher learning in experimental group (cool colors). Error metrics stay very close to zero indicating no difference between the two groups. We observe that Movements (repetitions 2, 3, and 4), Grasp Orientation Deviation (repetitions 2, 4, and 5) and Drive Path Deviation (In Plane) (repetition 2) show statistically significant higher performance improvement in the experimental group. Also, the statistically higher improvement in Movements for the experimental group in the TEACH mode repetition becomes smaller, and eventually, the control group improvement in Movements is higher in the FINAL repetition. Deficit metrics (lower four rows in Fig. 3) indicate higher learning in experimental group in the TEACH mode. This learning reduces compared to the control group by the FINAL repetition.

### Discussion and conclusion

Our automated teaching framework demonstrates the feasibility of an automated surgical coach that can deliver relevant, targeted, critical, and individualized learning. We implemented this framework onto a VR simulation platform in the context of needle passing in RAMIS. We demonstrated its effect on surgical technique improvement in a pilot randomized controlled trial. Our pilot study forms the basis for research and development of future automated teaching and coaching platforms in surgical skills training.

Our observation of higher learning in deficit metrics and lower improvement in motion efficiency among the experimental group compared to the control group is coherent with previous findings [3,6,7,23]. For example, Singh et al. [23] observed enhanced quality in performance at the expense of time and motion efficiency metrics while studying the effect of coaching on skill acquisition. Additionally, this is expected since the teaching cues are targeted at the CLE in needle passing to improve the surgical technique (deficit metrics) of the task. At the same time, motion efficiency metrics are meaningful only once the underlying technique has been mastered to deliver competent outcomes.

We exposed the experimental group to a single TEACH mode training session. As a result, the initial improvement in surgical technique during the TEACH session (reflected by the statistically significant improvement in deficit metrics in Fig. 3) became less significant by the FINAL session. A future study should include more number of task repetitions to effectively enable improvement in performance of trainees receiving the automated teaching intervention. We explored visual teaching cues in the current work. In the future, haptics-based cues that use virtual fixtures [4] can be added to provide a hand-over-hand guidance to demonstrate

### Table 1 Performance improvement from baseline on overall task execution at the final repetition

| Metric                               | Experimental (14) | Control (16) | $P$ value |
|--------------------------------------|-------------------|--------------|-----------|
| Completion time                       | $-132.71 (134.05)$| $-167.95 (172.90)$| 0.52      |
| Path length                           | $-137.04 (162.11)$| $-208.81 (227.47)$| 0.13      |
| Movements per second                  | 0.25 (0.39)       | 0.03 (0.45)  | 0.14      |
| Ribbon area                           | $-277.47 (322.11)$| $-427.91 (453.19)$| 0.15      |
| Console path length                   | $-337.42 (376.86)$| $-546.20 (428.33)$| 0.16      |
| Console workspace volume              | $-294.24 (528.96)$| $-882.42 (1403.51)$| 0.04      |
| Exc. needle tissue force (Count)      | $-1.50 (8.92)$    | $-5.88 (14.57)$| 0.14      |
| Exc. instrument force (Count)         | $-1.79 (2.86)$    | $-1.63 (5.80)$| 0.53      |
| Exc. instrument force (Time)          | $-6.35 (8.07)$    | $-5.72 (19.79)$| 0.47      |
| Exc. needle tissue force (Count)      | $-2.93 (5.51)$    | $-4.44 (15.59)$| 0.97      |
| Exc. needle tissue force (Time)       | $-11.81 (21.04)$  | $-17.70 (48.45)$| 0.76      |
| Grasp position dev. (Angle)           | $-3.73 (16.86)$   | $4.19 (18.32)$ | 0.31      |
| Grasp orientation dev. (Angle)        | $-14.53 (12.99)$  | $-4.22 (11.09)$| 0.04      |
| Drive path dev. (In Plane)            | $-0.00 (0.06)$    | $-0.01 (0.05)$ | 1.00      |
| Drive path dev. (Out of Plane)        | $-0.01 (0.05)$    | $-0.02 (0.03)$ | 1.00      |

Bold value indicates statistical significance ($p < 0.5$) Smaller values indicate higher improvement. $P$ values are for Mann–Whitney $U$ test. The numbers in the parentheses are standard deviations

Exc, excessive; dev., deviation, length in millimeters, time in seconds, angle in degrees
the critical learning elements and task segments that were subpar.

Our teaching framework is a step toward addressing current limitations of VR training, i.e., lack of accessible expert teaching and coaching. We address previous recommendations on development of automated tools to deliver expert-like feedback and coaching, since current approaches are limited due to scalability issues [23]. We chose to demonstrate our automated teaching framework in RAMIS VR simulation because it affords the best opportunity to provide surgeons with context-relevant feedback with minimal overhead. Technical skill acquired in VR transfers to benchtop simulation and OR [26].

Teaching and feedback are important components of effective surgical coaching [1,12,24]. Our framework realizes these components using teaching cues and deficit metrics. These cues and metrics identify the need for and offer immediate guidance to the learner on how to correct errors and reduce them by improving their technique. Future research should address the question of whether and to what extent, improvements in performance with automated surgical coaching transfer without attrition to the OR and eventually affect the safety and quality of patient care.

Acknowledgements We thank Anusha Balan, Sina Parastegari, Prasad, and Ashwin from Intuitive Surgical Inc. (ISI, Sunnyvale, California) for their help with the dVSim. We thank Simon DiMaio (ISI) for his feedback and helpful inputs. We thank Anthony Jarc (ISI) for help with the study IRB logistics. We thank Umut and Daniel from SenseGraphics AB (Kista, Sweden) for their support related to the H3DAPI.

Funding Anand Malpani was supported through a Link Foundation fellowship award for advanced simulation and training.

Compliance with ethical standards

Conflict of interest Anand Malpani was employed as an intern at Intuitive Surgical Inc. (Sunnyvale, California, USA) during part of this work. Henry C. Lin is an employee at Intuitive Surgical Inc. (Sunnyvale, California, USA). Intuitive Surgical Inc. manufactures the da Vinci Skills Simulator that was used as the platform to test the work on. Gregory D. Hager and Russell H. Taylor have received funding from Intuitive Surgical Inc. previously for other research projects. Johns Hopkins University and Intuitive Surgical Inc. have an ongoing partnership to support research on the surgical robots.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Western IRB, Protocol 20121049) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Becker AJ (2009) It’s not what they do, it’s how they do it: athlete experiences of great coaching. Int J Sports Sci Coach 4(1):93–119. https://doi.org/10.1260/1747-9541.4.1.93
2. Birkmeyer JD, Finks JF, O’Reilly A, Oerline M, Carlin AM, Nunn AR, Dimick J, Banerjee M, Birkmeyer NJ (2013) Surgical skill and complication rates after bariatric surgery. New Engl J Med 369(15):1434–1442. https://doi.org/10.1056/NEJMc1300625
