Surgery xxx (2019) 1–8

Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

The System for Telementoring with Augmented Reality (STAR): A head-mounted display to improve surgical coaching and confidence in remote areas

Edgar Rojas-Muñoz, BS[a], Maria E. Cabrera, PhD[b], Chengyuan Lin, BSc[c], Daniel Andersen, MS[c], Voicu Popescu, PhD[c], Kathryn Anderson, BSN, MSGH[d], Ben L. Zarzaur, MD, MPH[d], Brian Mullis, MD[d], Juan P. Wachs, PhD[a,*]

[a] School of Industrial Engineering, Purdue University, West Lafayette, IN
[b] Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA
[c] Department of Computer Science, Purdue University, West Lafayette, IN
[d] School of Medicine, Indiana University, Indianapolis, IN

ARTICLE INFO

Article history:
Accepted 3 November 2019
Available online xxx

ABSTRACT

Background: The surgical workforce particularly in rural regions needs novel approaches to reinforce the skills and confidence of health practitioners. Although conventional telementoring systems have proven beneficial to address this gap, the benefits of platforms of augmented reality-based telementoring in the coaching and confidence of medical personnel are yet to be evaluated.
Methods: A total of 20 participants were guided by remote expert surgeons to perform leg fasciotomies on cadavers under one of two conditions: (1) telementoring (with our System for Telementoring with Augmented Reality) or (2) independently reviewing the procedure beforehand. Using the Individual Performance Score and the Weighted Individual Performance Score, two on-site, expert surgeons evaluated the participants. Postexperiment metrics included number of errors, procedure completion time, and self-reported confidence scores. A total of six objective measurements were obtained to describe the self-reported confidence scores and the overall quality of the coaching. Additional analyses were performed based on the participants' expertise level.
Results: Participants using the System for Telementoring with Augmented Reality received 10% greater Weighted Individual Performance Score ($P = .03$) and performed 67% fewer errors ($P = .04$). Moreover, participants with lower surgical expertise that used the System for Telementoring with Augmented Reality received 17% greater Individual Performance Score ($P = .04$), 32% greater Weighted Individual Performance Score ($P < .01$) and performed 92% fewer errors ($P < .001$). In addition, participants using the System for Telementoring with Augmented Reality reported 25% more confidence in all evaluated aspects ($P < .03$). On average, participants using the System for Telementoring with Augmented Reality received augmented reality guidance 19 times on average and received guidance for 47% of their total task completion time.
Conclusion: Participants using the System for Telementoring with Augmented Reality performed leg fasciotomies with fewer errors and received better performance scores. In addition, participants using the System for Telementoring with Augmented Reality reported being more confident when performing fasciotomies under telementoring. Augmented Reality Head-Mounted Display–based telementoring successfully provided confidence and coaching to medical personnel.

Introduction

Several studies project a decrease in the surgical workforce, particularly in rural and low- and middle-income regions.[1–3] This trend is attributable in part to the limited training opportunities for rural surgeons.[4–6] Surgery coaching programs, wherein one-on-one

* Reprint requests: Juan P. Wachs, PhD, Purdue University, 315 N. Grant Street, West Lafayette, IN 47907.
E-mail address: jpwachs@purdue.edu (J.P. Wachs).

https://doi.org/10.1016/j.surg.2019.11.008
0039-6060/

Fig 1. Participants are assigned randomly to an experimental condition. Participant receiving remote instruction using our ARHMD-based system (*right*). Participant receiving no external guidance beyond initial consultation of a medical course manual (*left*).

training sessions are held with a coaching surgeon who evaluates and corrects the performance of less-experienced surgeons, are important training tools that have been used to alleviate these limited training opportunities.[7,8] These approaches, however, are not scalable. Surgeons must be physically present at these coaching sessions, which imposes considerable constraints on time and budgets.

The advent of telementoring systems has alleviated such constraints related to physical presence without introducing greater operative times or complication rates.[9,10] In telementored surgical coaching, nonspecialist surgeons (mentees) receive operative guidance and coaching from remotely located specialists (mentors),[11–13] allowing operative innovations to be disseminated more rapidly.[14,15] Several studies have demonstrated the benefits of telementoring and video-based coaching in assessing remotely the technical skills of surgeons,[12,16] and providing assistance in rural settings.[17,18]

Augmented reality (AR) technologies are being integrated currently into telementoring systems.[19,20] Using AR, the local mentees can visualize expert-authored operative instructions directly in their field of view, as opposed to traditional telementoring systems that require mentees to shift focus away of the operating field to visualize the instructions.[21] These AR telementoring systems, however, are still in exploratory stages, and no studies have been performed to assess the effectiveness of such coaching using these platforms. For example, most studies that have assessed how to leverage telementoring systems to perform coaching have been focused on the impact of live video feedback between mentors and mentees.[22] However, the effectiveness of AR-generated guidance to perform coaching, such as virtual models of surgical instruments and incisions lines, is yet to be evaluated.

To address this gap, this study presents an evaluation of the effectiveness of such coaching using the System for Telementoring with Augmented Reality (STAR). STAR is a novel platform that leverages an AR, head-mounted display (ARHMD) worn by the mentee surgeon to display mentor-authored operative instructions. Mentees wearing the ARHMD can visualize these expert instructions as three-dimensional (3-D) overlays directly onto their field of view of the patient's body. To analyze the potential of STAR to provide confidence and coaching to medical personnel, this report presents an analysis of the coaching given by expert surgeons as they mentored medical personnel remotely through a leg fasciotomy training procedure on cadaveric specimens.

Methods

Ethical approval (IRB #1409037680) was obtained from Purdue University and Indiana University School of Medicine, and written participant consent was acquired for each participant. Twenty participants performed a leg fasciotomy training session on cadaveric specimens under one of two conditions: (1) receiving remote instruction using our ARHMD system (STAR, Fig 1, *right*) and (2) receiving no external guidance beyond initial consultation of the Advanced Surgical Skills for Exposure in Trauma course manual (control, Fig 1, *left*). The control condition is a placeholder for a surgeon with limited experience who may be requested to proceed with a procedure in a rural or austere environment. The experiment tested the procedural and confidence outcomes between participants receiving remote guidance with STAR against participants that only had access to written mentoring material.

Equipment and setting

STAR is a surgical telementoring platform that leverages an ARHMD to display operative instructions directly into the field of view of a mentee requiring assistance.[23,24] Figure 2 provides an artist's rendition of our clinical validation setup. At the mentee site, a top-down camera captures a live video feed of the operative field, which is sent to the remote mentor (Fig 2, *A*). This feed is displayed on a large interactive display (Fig 2, *B*). The remote mentor uses touch-based interactions to create technical annotations (Fig 2, *C*) over this video feed. These annotations are sent back to the ARHMD (Fig 2, *D*) and appear as 3-D imagery superimposed onto the mentee's field of view of the operative field (Fig 2, E). These annotations can provide the local mentee with coaching from the remote mentor (Fig 3).

The experiment was conducted at two separate facilities (approximately 1,600 feet away) at the campus of Indiana University School of Medicine. The remote mentor's setup included the large interactive display (Aquos Board PN-L603B, Sharp Electronics, Osaka, Japan) and a conferencing speakerphone (Konftel 300Mx,

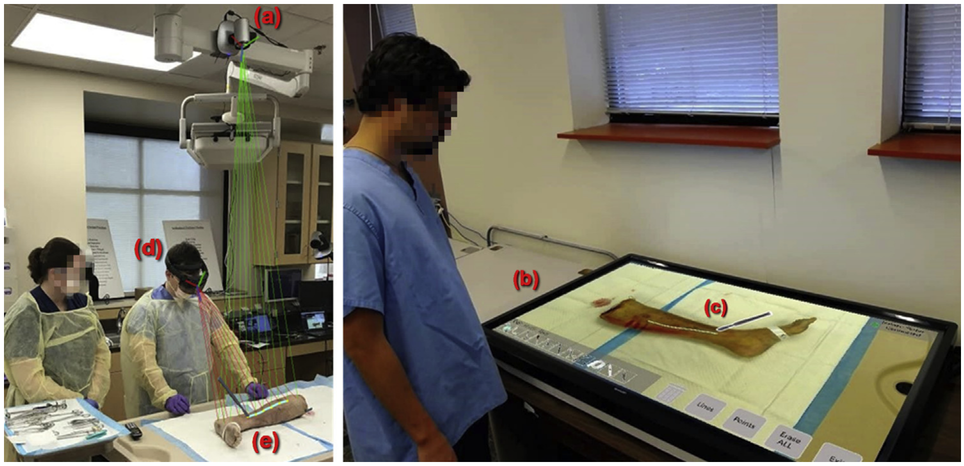

Fig 2. Artist rendition of the STAR platform. A live video feed of the operative field is acquired with a camera (a) and sent to the remote mentor. The remote mentor visualizes the video on a display (b), and creates operative instructions over this image using multi-touch interactions (c). Using our ARHMD platform (d), the mentee can visualize these annotations superimposed onto the patient's body, in 3-D (e).

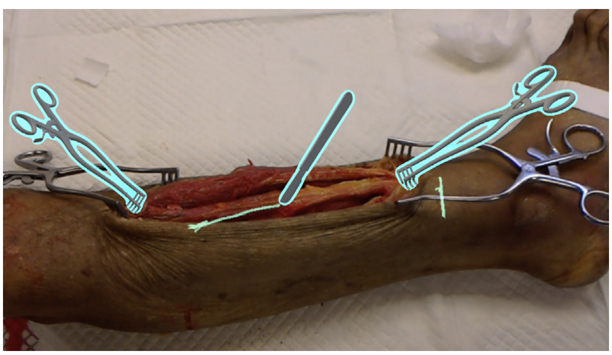

Fig 3. Examples of Augmented Reality 3-D annotations, visualized from the mentee's perspective. Thanks to the internal routines of the ARHMD, the operative instructions are projected onto the patient's body at the correct position and depth.

Konftel AB, Umeå, Sweden), providing audio communication with the mentee. Two stations were established at the mentee site, each equipped with one operating table, surgical lights, and a Mayo tray. One nurse assistant was stationed at each station and handed instruments to the mentees throughout the procedure. In addition, the STAR station included a top-down, pan/zoom/tilt camera (PTZ Pro 2, Logitech, Lausanne, Switzerland) attached to the surgical lights, the ARHMD (Microsoft HoloLens, Microsoft Corporation, Redmond, WA, USA) worn by the mentee, and a phone providing audio communication with the remote mentor.

The use of a Microsoft HoloLens as our ARHMD device makes our platform portable and untethered. All the required processing and rendering is done by the on-board computer of the device, which is totally wireless. We envision our system as an affordable and accessible approach to provide operative coaching. The system requires a one-time purchase of the ARHMD device (usually less than US$3,000) at the mentee site and a computer with touch-display based capabilities at the mentor site (usually less than US$1,000). Our research team created two, standalone apps, one running in the ARHMD at the mentee site and another running in the computer at the mentor site. These apps are available in GitHub (Mentor System App: www.github.com/edkazar/MentorSystemUWPWebRTC/releases; Mentee System App: www.github.com/practisebody/STAR/releases) and can be installed in the devices by following a simple installation guide. In addition, the systems connect through the internet using WebRTC, a video transmission protocol that automatically adapts the quality of the video based on the available bandwidth of the network (similar to the one used by Skype).

Participants

The population was composed of surgery residents and medical students. These groups were expected to have relatively less expertise, making this an ideal setup to compare whether the external sources of mentoring were producing benefits in the performance of the participants.[15,25] To investigate the effect of the mentoring conditions with respect to the participant's expertise, participants were recruited from three different strata to encompass various expertise levels: medical students, residents with only 1 year of postgraduate expertise and residents with 2 years or more of postgraduate expertise.

Metrics

Individual Performance Score and Weighted Individual Performance Score

The technical and nontechnical skills of the participants were evaluated on-site, using the Individual Performance Score (IPS),[26] adapted for fasciotomy procedures by expert surgeons of the team (Supplemental Material 1). The evaluation includes six sections:

- Anatomic landmarks (A)—to be highlighted before any incision is performed.
- Anterolateral incision (B)—where the participants needed to identify and release the fascia from the anterior and lateral compartments.
- Posteromedial incision (C)—where participants were required to identify and release the superficial posterior and the deep posterior compartments.
- General performance evaluation included three sections: technique points (D), operative field maneuvers (E), and instrument use (F).

IPS, however, assumes each of these six sections to be equally important. A weighted IPS (WIPS) was created to relax this assumption and obtain a more comprehensive performance score.

A survey was conducted with nine experienced, general and orthopedic trauma attending surgeons to determine an importance of the IPS sections. The survey revealed the following relationship: $F < D < E < C = B < A$ (eg, instrument use was determined to be the least important section, and identification of anatomic landmarks was determined to be the most important section). This unequal importance is represented with normalized numeric weights: A → 0.236, B → 0.217, C → 0.217, D → 0.106, E → 0.162, and F → 0.062. These weights were used to normalize the IPS values, leading to a comprehensive score that considers the unequal importance of the procedural aspects.

The procedure completion time and the number of errors during the procedure (out of 11 possible errors highlighted as part of IPS) were also analyzed as postexperiment metrics.

Self-reported confidence score

A self-reported confidence score quantified the participants' confidence in performing leg fasciotomies before and after the experiment. Participants filled a questionnaire composed of four, five-level Likert scale questions that evaluated the participants' knowledge on anatomic landmarks and procedural steps, instrument handling technique, and confidence to perform the procedure independently.

Quantifications of confidence and coaching

Six additional measurements were obtained to describe objectively the confidence scores and the overall quality of coaching. These measurements were: (1) the number of operative AR annotations created by the remote mentor; (2) the number of times the mentor asked the mentee for confirmation (eg, "*That structure looks like the nerve to me, do you think the same?*"); (3) the number of times the mentee asked the mentor for instruction (eg, "*There is muscle at the posterior border of the tibia, what would you like me to do?*"); (4) the number of times the mentee asked the mentor for confirmation (eg, "*Are you sure I can cut this?*"); (5) the number of corrections given by the mentor (eg, "*No, use your scissors for that, not the knife*"); and (6) the percentage of the total completion time during which the mentee received guidance from the remote mentor. These measurements were obtained only for participants in the STAR condition, as participants in the control condition did not receive telementoring.

Statistical analysis

Our statistical analysis treated the experimental conditions as independent variables and all the metrics as dependent variables. The null hypothesis for all comparisons was that both conditions (control and STAR) were the same. For the alternative hypothesis, all metrics were considered for a one-sided comparison: STAR less than control for errors and completion time, and STAR greater than control for all other metrics. The normality assumption of the data was evaluated using the Shapiro-Wilk test.[27] When data pointed to non-normality, the nonparametric Mann-Whitney *U* test was used to compare populations with unpaired data; whereas the nonparametric Wilcoxon signed-rank test was used to compare populations with paired data.[28,29] For normal data, a Levene's test was used to assess the equal variance condition of the data.[30] For data with equal variance, a 2-sample *t*-test was used for comparisons.[30]

Results

A total of 14 surgery residents and 6 medical students (age mean 28.5 years ± 3.3 years; 13 males, 7 female) were distributed equally between the 2 experimental conditions for a between-subjects design ($n = 10$). Medical students varied between their first year (2), second year (1) and fourth year (3) of training. Surgery residents were in their first year (4), second year (3), third year (3), or fourth year (4) of training in general and orthopedic surgery. Based on the expertise-based strata, 3 expertise-based subgroups were considered for analyses: only medical students ($n = 6$), only residents ($n = 14$), and a combination of medical students and first-year surgery residents ($n = 10$). The latter was considered the subgroup who would benefit the most from the coaching experience because of their relatively lesser expertise.

Expert attending surgeons assessed the onsite\performance of the participants. Table I summarizes the results for all performance-based metrics. STAR participants received 10% greater WIPS ($P = .03$) and made 67% fewer errors ($P = .04$). No statistically significant difference was found between task completion times. Table II reports the scores for the 3 subgroups (only residents, only medical students, and medical students and first-year residents combined), in terms of IPS, WIPS, and number of errors. The null hypothesis could not be rejected in any of the metrics when considering only residents. When considering only medical students, STAR participants received 22% greater WIPS ($P = .05$) and made 89% fewer errors ($P = .02$). Finally, when considering the low-expertise subgroup, STAR participants received 17% greater IPS ($P = .04$), 32% greater WIPS ($P < .01$), and made 92% fewer errors ($P < .001$).

Table III reports the participants' confidence scores for both experimental conditions, before and after the experiment. STAR participants reported 25% more confidence in all evaluated aspects ($P < .03$). We did not find statistical significance in the confidence scores in the control condition ($P > .09$). In addition, Table IV reports the measurements of confidence and coaching for participants in the STAR condition, divided by the 3 population subgroups. Medical students received more ($P = .04$) corrections compared with residents.

Table V presents the distribution of the errors highlighted by IPS with respect to the mentoring condition for the different, expertise-based subgroups. The errors were divided into different levels of occurrence based on how many participants from each of the expertise-based subgroups incurred in such error. The error was classified as low frequency if less than 20% of the participants in the subgroup incurred into the error. The error was classified as medium frequency if between 20% and 40% of the participants in the subgroup incurred into the error. Finally, the error was classified as high frequency if more than 40% of the participants in the subgroup incurred into the error. According to our scheme of low-medium-high frequency classification, the subgroup with only medical students had nine errors classified as low frequency, one classified as medium frequency, and one classified as high frequency; the subgroup with only residents had eight errors classified as low frequency, three classified as medium frequency, and none classified as high frequency; and the subgroup with medical students and first-year residents combined had six errors classified as low frequency, five classified as medium frequency, and none classified as high frequency.

A 2-sample *t*-test was used to evaluate the hypothesis of whether receiving mentoring using the STAR condition decreased the occurrence of errors was performed. The results revealed that participants in the STAR condition performed less low-frequency errors ($P = .05$), as well as less medium-frequency errors ($P < .001$). No statistical analyses were run for high-frequency errors, because not enough errors were classified into this category.

Discussion

Our results follow the trends of earlier telementoring studies in that operative outcomes can be increased when remote specialists guide the local mentees through a procedure.[19] Our hypothesis,

Table I
Summary of results for all quantitative metrics averaged over participants in each condition

Metrics	STAR	Control	*P* value
Completion time, median (95% CI), min	22.98 (12.08–46.00)	24.60 (9.35–32.33)	.6
IPS, mean (95% CI)	483.1 (436.4–529.8)	437.4 (368.2–506.7)	.1
WIPS, mean (95% CI), %	82.70 (77.17–88.13)	74.60 (63.96–85.20)	.03*
Errors, mean (95% CI), count	0.6 (0.1–1.1)	1.8 (0.8–2.8)	.04*

* A significant difference between the metrics.

Table II
Participants' performance considering the expertise subgroups

Metric	STAR			Control		
	Medical students ($n = 3$)	Residents ($n = 7$)	Low expertise ($n = 5$)	Medical students ($n = 3$)	Residents ($n = 7$)	Low expertise ($n = 5$)
IPS, mean (95% CI)	462.8 (242.7–682.8)	491.8 (437.2–546.4)	451.5 (365.0–538.1)	366.5 (220.4–512.6)	467.8 (378.6–557.0)	376.4 (319.0–433.8)
WIPS, mean (95% CI),%	81.97 (56.24–107.69)	82.94 (76.28–89.60)	80.50 (70.69–90.31)	64.33 (37.15–91.51)	78.97 (65.4–92.55)	54.92 (45.26–74.58)
Errors, mean (95% CI), count	0.33 (0.01–1.86)	0.71 (0.22–1.30)	0.20 (0.01–1.11)	2.67 (1.15–5.25)	1.43 (1.07–2.84)	2.60 (1.38–4.45)

Table III
Participants' self-reported confidence scores

Confidence Assessment Aspect	STAR			Control		
	Pre SRCS, median (95% CI)	Post SRCS, median (95% CI)	*P* value	Pre SRCS, median (95% CI)	Post SRCS, median (95% CI)	*P* value
Identify anatomical landmarks	3.0 (1.0–4.0)	4.0 (3.0–5.0)	.019*	3.5 (2.0–5.0)	4.0 (3.0–5.0)	.082
Knowledge of procedural steps	3.0 (1.0–3.0)	4.0 (3.0–5.0)	<.001*	2.5 (1.0–5.0)	3.5 (3.0–5.0)	.089
Instrument handling technique	3.0 (2.0–5.0)	4.5 (3.0–5.0)	.034*	4.0 (2.0–5.0)	4.0 (1.0–5.0)	.473
Perform procedure alone	2.0 (1.0–3.0)	3.5 (2.0–4.0)	.002*	3.0 (1.0–5.0)	3.5 (2.0–5.0)	.199

SRCS, self-reported confidence score.
* A statistically significant improvement in the participant's confidence level.

nonetheless, evaluated whether mentees using the STAR ARHMD-based telementoring approach could perform a leg fasciotomy more quickly, better, and with fewer mistakes. Our hypothesis also evaluated the confidence of the participants in performing the fasciotomies.

STAR participants made fewer errors when performing the procedure. This is a critical outcome, because it identifies and avoids operative errors that are critical for patient safety and also predicts technical skills and performance.[31,32] STAR participants also achieved greater technical and nontechnical skills scores according to IPS, and significantly greater technical and nontechnical skills scores according to WIPS.

The expertise-based subgroups revealed surprising trends. Medical students in the STAR condition performed fewer errors and obtained greater WIPS compared with participants in the control condition. Although the size of this subgroup was low, the results suggest that participants with low expertise performed the fasciotomies in an overall better fashion when using the STAR platform. The medical students and first-year resident subgroup also revealed this same trend. Participants in the STAR condition performed fewer errors and obtained greater IPS and WIPS compared with participants in the control condition. In contrast, the residents subgroup did not report statistically significant differences between the experimental conditions. This observation hints that surgery residents, particularly those with more years of postgraduate experience, found telementored guidance through a leg fasciotomy to be irrelevant for their performance. For example, one experienced resident was able to complete the procedure without relying on the external guidance and obtaining a near-perfect score. We hypothesize that the relatively low difficulty of the operative procedure caused this last trend. Therefore, future validations should be performed with more complex operations that may reveal the benefits of telementoring for health practitioners with greater levels of expertise levels.

STAR participants reported marked improvements in all evaluated aspects of their confidence scores. These results demonstrate that an interactive, telementoring experience with the STAR ARHMD had a positive impact in the confidence of the participants. Although studies have shown that health practitioners' confidence in their surgical skills is correlated to competence and self-assessment of their skill,[33,34] surveys report that the health practitioners' confidence in their skills is not particularly high.[35] Extrapolating our results, the integration of STAR and similar telementoring platforms to current coaching programs could help reinforcing surgical knowledge and enhancing the self-confidence of health practitioners.

The measurements of confidence and coaching elaborated on the self-reported confidence scores. The remote mentors created 19 annotations per participant on average. The use of the annotations can be divided into 4 situations: (1) exemplifying which instrument to use (eg, placing the icon of a scalpel after saying "*Cut here*"); (2) showing the location of anatomic structures (eg, drawing a circle around the peroneal nerve); (3) showing the length and location of incisions (eg, drawing a line along the leg to depict where to cut); and (4) acquiring a better awareness of the operating field (eg, drawing a circle around the toe to determine the orientation of the leg). By creating these annotations, the mentor was able to convey more guidance, which may be a possible reason of the increased performance and confidence scores of STAR participants.

Moreover as described by Green et al,[22] transmitting the real-time visual feedback of the operating field allowed the remote mentor to provide better coaching. The visual feedback allowed the mentor to ask for confirmation 7 times per participant on average

Table IV
Quantifications of coaching and confidence for the different expertise-based subgroups

Measurements of coaching and confidence	Medical students (*N* = 3)	Residents (*N* = 7)	Low expertise (*N* = 5)
Number of AR annotations created, mean (95% CI), count	18.00 (13.52–23.49)	19.29 (16.17–22.83)	20.00 (16.27–24.33)
Number of times the mentor asked for confirmation, mean (95% CI), count	8.67 (5.66–12.70)	6.14 (4.45–8.27)	8.00 (5.72–10.89)
Number of times the mentee asked for instruction, mean (95% CI), count	5.67 (3.30–9.07)	4.29 (2.89–6.12)	5.20 (3.40–7.62)
Number of times the mentee asked for confirmation, mean (95% CI), count	3.00 (1.37–5.69)	5.29 (3.72–7.29)	4.20 (2.60–6.42)
Number of corrections given by the mentor, mean (95% CI), count	7.00 (4.33–10.70)	3.86 (2.54–5.61)	6.40 (4.38–9.03)
Time during which the mentee received guidance, mean (95% CI), percentage	44.2 (39.5–48.9)	49.15 (42.78–55.52)	44.51 (42.33–46.69)

Table V
Distribution of the errors highlighted by IPS with respect to the mentoring condition for the different expertise-based subgroups

Error code	Error description	Medical students (*N* = 6)		Residents (*N* = 14)		Low expertise (*N* = 10)	
		STAR (*n* = 3)	CONTROL (*n* = 3)	STAR (*n* = 7)	CONTROL (*n* = 7)	STAR (*n* = 5)	CONTROL (*n* = 5)
E1	After initial anterolateral incision, uses scalpel or scissor instead of blunt instrument to avoid damaging the superficial peroneal nerve	0	0	1	2	0	1
		Low frequency		Medium frequency		Low frequency	
E2	Does not protect superficial peroneal nerve while extending the intermuscular septum incision over the lateral and anterior compartments	0	1	1	3	0	2
		Low frequency		Medium frequency		Medium frequency	
E3	Incorrectly identifies and releases the anterior compartment	0	0	0	0	0	0
		Low frequency		Low frequency		Low frequency	
E4	The tip of the scissors was not directed away from the intermuscular septum while releasing the anterior compartment	0	1	0	1	0	2
		Low frequency		Low frequency		Medium frequency	
E5	Incorrectly identifies and releases the lateral compartment	0	1	0	0	0	1
		Low frequency		Low frequency		Low frequency	
E6	The tip of the scissors was not directed away from the intermuscular septum while releasing the lateral compartment	0	1	0	3	0	2
		Low frequency		Medium frequency		Medium frequency	
E7	After initial posteromedial incision, uses scalpel or scissor instead of blunt instrument to avoid damaging the superficial peroneal nerve	0	0	0	0	0	0
		Low frequency		Low frequency		Low frequency	
E8	Does not identify and retract the saphenous vein posteriorly	1	2	1	0	1	2
		High frequency		Low frequency		Medium frequency	
E9	Incorrectly identifies and releases the superficial posterior compartment	0	0	1	0	0	0
		Low frequency		Low frequency		Low frequency	
E10	Incorrectly identifies and releases the deep posterior compartment	0	2	1	1	0	3
		Medium frequency		Low frequency		Medium frequency	
E11	Fails to protect the neurovascular bundle while releasing the deep posterior compartment	0	0	0	0	0	0
		Low frequency		Low frequency		Low frequency	

and to perform 5 corrections per participants on average. The following transcription exemplifies one of these situations:

Mentor: "That looks like the saphenous vein."
Mentee: "What should I do with it?"
Mentor: "Continue with your incision, just make sure to stay away from the vein."
Mentee continues with the incision and gets dangerously close to the vein.
Mentor: "Wow! Be careful there, you almost got the vein in the last movement you did. Try not to get your knife too close to this area (the mentor draws a circle in the screen)*."*

In this example, the mentor corrected the mentee and provided more details about which area to avoid thanks to the visual feedback and the AR annotations. The visual feedback also allowed the mentee to ask for instructions and confirmations an average of 5 times per procedure, as depicted in the following example:

Mentee: There is still some muscle here (the mentee points at a specific point in the leg). *I think I should cut there more.*
Mentor: No, it is okay if there is still some muscle there.

Finally, the percentage of time during which mentees received guidance from the remote mentor can be associated with the increased performance and increased confidence scores. On average, the mentees received guidance for 10 minutes and 48 seconds, which represented 47% of the total time to complete the task. These results suggest that the STAR participants received remote guidance for almost half of the time they took to complete the task without incurring increasing completion times.

Table V presents a breakdown of the errors highlighted by IPS for the different expertise-based subgroups. Breaking down the errors in this way allowed us to analyze which steps were more difficult and whether receiving mentoring using the STAR platform decreased the amount of times each specific error was performed. Based on our low-medium-high frequency classification scheme, errors E3, E5, E7, E9, and E11 were classified as low frequency for all expertise-based subgroups. Most of these possible errors were related to incorrectly identifying and releasing the compartments (anterior, lateral, and posterior). These results show that participants were able to release the outermost leg compartments without difficulties. Nonetheless, the process of releasing the deep posterior compartment (E10) was considered as medium frequency for both the subgroup of only medical students and the subgroup of medical students and first-year residents. These results are not unexpected, because the process of releasing the deep posterior compartment can be more error prone. Therefore, considering these results, we recommended paying extra attention to the instruction process of releasing this compartment while performing training for fasciotomies. Errors related to how participants used their tools (E1, E4, E6, and E7) were considered of medium frequency in most the cases (except E7). This observation reveals a deficiency in the way participants handled their operative instruments. As a result, special emphasis should be placed in the correct use of operative instruments during the residency years of future surgical personnel. Errors related to the identification and protection of anatomic landmarks (E2, E8, E7, and E11) were the most often incurred errors during our experiment, to the point of being considered of high frequency for the subgroup of only medical students. This revealed an aspect that should be reinforced in current residency programs, because the ability of identifying and protecting internal anatomic structures is critical to proper operative performance. These insights and the results from the statistical analyses reaffirm the validity of our platform as a novel method to increase medical confidence and provide accessible coaching.

Regarding limitations, participants mentioned having greater-than-normal levels of anxiety during the experiment attributable to the authority position of their evaluators and mentors (ie, senior faculty members from their surgery residency). In addition, limitations in the 3-D tracking of the ARHMD caused the virtual annotations to appear to drift in space (approximately 1 cm) from the mentee's perspective of the operating field. These problems, however, are known limitations of current ARHMD systems. Moreover, some STAR participants stated that the weight of the ARHMD interfered with their posture and comfort. This limitation, however, was not shared by the on-site, expert evaluators. These expert evaluators wore the device and commented that our telementoring platform was lighter than the headlamps and/or helmets with fans that orthopedic surgeons routinely wear for total joint arthroplasty, a procedure that usually last from 60 to 120 minutes.

Venues for future work include the following possibilities: First, our system should be evaluated against on-site mentoring to obtain objective and subjective comparisons. Although on-site mentoring is the standard approach, telementoring systems provide advantages for rural medicine or combat medicine where an expert is not on site. For example, several studies have concluded that telementoring can be as effective as local mentoring to develop operative skills.[13,36] Nonetheless, comparisons between these approaches in terms of workload, stress, and frustration are yet to be explored. Future studies should be performed to evaluate our system in situations with greater ecologic validity (eg, severe bleeding from actual patients). Although we do not anticipate that such situations will have a negative impact in the telementoring capabilities of our platform, more experiments should be conducted. In contrast, we anticipate our system to be more beneficial in these settings than conventional telementoring systems. For example, our platform will allow the remote mentor to provide real-time guidance based on the visual feedback provided by the system, which can be critical in situations when errors arise. In addition, the transition of our platform into a fully commercial product remains unaddressed. In addition to obtaining US Food and Drug Administration approval, the legal implications of such a platform need to be defined and addressed before its integration into surgical curricula. For example, hospital and insurance companies should be consulted to create a model of how to integrate ARHMD telementoring platforms into hospital coaching programs.

In conclusion, this work evaluated whether the novel telementoring platform referred to as STAR based on an ARHMD could be used for more effective coaching of surgical personnel in remote locations. Participants using our ARHMD-based telementoring platform performed leg fasciotomies in less time, with fewer errors, and receiving better procedural scores. In addition, participants experienced greater increases in their confidence thanks to the visual feedback of the platform and the AR operative instructions. Our work suggests that ARHMD-based telementoring may be an effective approach for low-cost and accessible coaching and training.

Funding/Support

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs, under Award Number W81XWH-14-1-0042, and by the National Science Foundation, under Grant DGE-1333468. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the funders.

Conflict of interest/Disclosure

Voicu Popescu and Juan Pablo Wachs are inventors on a US-issued patent submitted by Purdue Research Foundation, which is relevant to this work. In addition, all authors received funding from the grant reported in the Funding/Support section of this report.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.surg.2019.11.008.

References

1. Hoyler M, Finlayson SR, McClain CD, Meara JG, Hagander L. Shortage of doctors, shortage of data: a review of the global surgery, obstetrics, and anesthesia workforce literature. *World J Surg*. 2014;38:269–280.
2. Thompson MJ, Lynge DC, Larson EH, Tachawachira P, Hart LG. Characterizing the general surgery workforce in rural America. *Arch Surg*. 2005;140:74–79.
3. Goodman DC. Twenty-year trends in regional variations in the US physician workforce: variations in physician supply have persisted for decades, but their effects on Americans' health are still being investigated. *Health Aff (Millwood)*. 2004;23(Suppl Variation):VAR90–VAR97.
4. Doty B, Zuckerman R. Rural surgery: framing the issues. *Surg Clin*. 2009;89:1279–1284.
5. Burkholder HC, Cofer JB. Rural surgery training: a survey of program directors. *J Am Coll Surg*. 2007;204:416–421.
6. Doty B, Zuckerman R, Borgstrom D. Are general surgery residency programs likely to prepare future rural surgeons? *J Surg Educ*. 2009;66:74–79.
7. Greenberg CC, Ghousseini HN, Quamme SRP, Beasley HL, Wiegmann DA. Surgical coaching for individual performance improvement. *Ann Surg*. 2015;261:32–34.
8. Stefanidis D, Anderson-Montoya B, Higgins RV, et al. Developing a coaching mechanism for practicing surgeons. *Surgery*. 2016;160:536–545.
9. Hung AJ, Chen J, Shah A, Gill IS. Telementoring and telesurgery for minimally invasive procedures. *J Urol*. 2017;199:355–369.
10. Bilgic E, Turkdogan S, Watanabe Y, et al. Effectiveness of telementoring in surgery compared with on-site mentoring: a systematic review. *Surg Innov*. 2017;24:379–385.
11. Boedeker D, Numanoglu A, Hall T, Boedeker B. Description of a novel telemedicine imaging system which could be used to teach surgical burn management techniques by telementoring in remote areas. *J Int Soc Telemed EHealth*. 2017;5:GKR-e12.
12. Greenberg CC, Ghousseini HN, Quamme SRP, et al. A statewide surgical coaching program provides opportunity for continuous professional development. *Ann Surg*. 2018;267:868–873.
13. Erridge S, Yeung DK, Patel HR, Purkayastha S. Telementoring of surgeons: a systematic review. *Surg Innov*. 2019;26:95–111.
14. Treter S, Perrier N, Sosa JA, Roman S. Telementoring: a multi-institutional experience with the introduction of a novel surgical approach for adrenalectomy. *Ann Surg Oncol*. 2013;20:2754–2758.
15. Snyderman CH, Gardner PA, Lanisnik B, Ravnik J. Surgical telementoring: a new model for surgical training. *Laryngoscope*. 2016;126:1334–1338.
16. Greenberg CC, Dombrowski J, Dimick JB. Video-based surgical coaching: an emerging approach to performance improvement. *JAMA Surg*. 2016;151:282–283.
17. Sebajang H, Trudeau P, Dougall A, Hegge S, McKinley C, Anvari M. Telementoring: an important enabling tool for the community surgeon. *Surg Innov*. 2005;12:327–331.
18. Sebajang H, Trudeau P, Dougall A, Hegge S, McKinley C, Anvari M. The role of telementoring and telerobotic assistance in the provision of laparoscopic colorectal surgery in rural areas. *Surg Endosc Interv Tech*. 2006;20:1389–1393.
19. Talbot M, Harvey E, Berry G, et al. A pilot study of surgical telementoring for leg fasciotomy. *J R Army Med Corps*. 2018;164:83–86.
20. Andersen D, Popescu V, Cabrera ME, et al. Medical telementoring using an augmented reality transparent display. *Surgery*. 2016;159:1646–1653.

21. Budrionis A, Augestad KM, Patel HR, Bellika JG. An evaluation framework for defining the contributions of telestration in surgical telementoring. *Interact J Med Res*. 2013;2:e14.
22. Green JL, Suresh V, Bittar P, Ledbetter L, Mithani SK, Allori A. The utilization of video technology in surgical education: a systematic review. *J Surg Res*. 2019;235:171–180.
23. Lin C, Andersen D, Popescu V, et al. A first-person mentee second-person mentor AR interface for surgical telementoring. In: *IEEE 2018 international symposium on mixed and augmented reality*. Piscataway, NJ: IEEE; 2018: 3–8.
24. Rojas-Muñoz E, Cabrera ME, Andersen D, et al. Surgical telementoring without encumbrance: a comparative study of see-through augmented reality-based approaches. *Ann Surg*. 2018;270:384–389.
25. Rafiq A, Moore JA, Zhao X, Doarn CR, Merrell RC. Digital video capture and synchronous consultation in open surgery. *Ann Surg*. 2004;239: 567–573.
26. Mackenzie CF, Garofalo E, Shackelford S, et al. Using an individual procedure score before and after the advanced surgical skills exposure for trauma course training to benchmark a hemorrhage-control performance metric. *J Surg Educ*. 2015;72:1278–1289.
27. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). *Biometrika*. 1965;52:591–611.
28. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18:50–60.
29. Wilcoxon F. Individual comparisons by ranking methods. *Biom Bull*. 1945;1: 80–83.
30. Levene H. Robust tests for equality of variances. *Contrib Probab Stat Essays Honor Harold Hotel*. 1961:279–292.
31. Bann S, Khan M, Datta V, Darzi A. Surgical skill is predicted by the ability to detect errors. *Am J Surg*. 2005;189:412–415.
32. Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, Sundt 3rd TM. Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. *Surgery*. 2007;142:658–665.
33. Leopold SS, Morgan HD, Kadel NJ, Gardner GC, Schaad DC, Wolf FM. Impact of educational intervention on confidence and competence in the performance of a simple surgical task. *J Bone Joint Surg Am*. 2005;87:1031–1037.
34. Clanton J, Gardner A, Cheung M, Mellert L, Evancho-Chapman M, George RL. The relationship between confidence and competence in the development of surgical skills. *J Surg Educ*. 2014;71:405–412.
35. Bucholz EM, Sue GR, Yeo H, Roman SA, Bell RH, Sosa JA. Our trainees' confidence: results from a national survey of 4136 US general surgery residents. *Arch Surg*. 2011;146, 907-114.
36. Huang EY, Knight S, Guetter CR, et al. Telemedicine and telementoring in the surgical specialties: a narrative review. *Am J Surg*. 2019;218:760–766.