Published on Journal of Medical Robotics Research, this is a Pre-print

# Evaluation of Pre-Training with the da Vinci Skills Simulator on Motor Skill Acquisition in a Surgical Robotics Curriculum

Edoardo Battaglia[a], Bradly Mueller[b], Deborah Hogg[b], Robert Rege[b], Daniel Scott[b], and Ann Majewicz Fey[a]

[a] *Department of Mechanical Engineering, the University of Texas at Austin, TX 78712 USA*
*e-mail: edoardo.battaglia@utexas.edu*

[b] *Simulation Center and Department of Surgery, the University of Texas Southwestern Medical Center, TX 75390 USA*

Training for robotic surgery can be challenging due the complexity of the technology, as well as a high demand for the robotic systems that must be primarily used for clinical care. While robotic surgical skills are traditionally trained using the robotic hardware coupled with physical simulated tissue models and test-beds, there has been an increasing interest in using virtual reality simulators. Use of virtual reality (VR) comes with some advantages, such as the ability to record and track metrics associated with learning. However, evidence of skill transfer from virtual environments to physical robotic tasks has yet to be fully demonstrated. In this work, we evaluate the effect of virtual reality pre-training on performance during a standardized robotic dry-lab training curriculum, where trainees perform a set of tasks and are evaluated with a score based on completion time and errors made during the task. Results show that virtual reality pre-training is weakly significant (p $\leq$0.1) in reducing the number of repetitions required to achieve proficiency on the robotic task; however, it is not able to significantly improve performance in any robotic tasks. This suggests that important skills are learned during physical training with the surgical robotic system that cannot yet be replaced with virtual reality training.

*Keywords*: Training; Virtual Reality; Robotic Surgery.

## 1. Introduction

Robotic surgery has become a concrete reality in clinical practice and is used routinely in a variety of medical procedures [1]. Indeed, the years between 2012 and 2018 saw a large increase in adoption of robotic surgery for all general surgery procedures from 1.8% to 15.1% [2]. While there is still no consensus in the scientific community on whether surgical robotics is inherently superior to traditional surgery, some studies highlight better morbidity rates [3] and others show at least no significant difference in outcome between robotic and traditional surgery [4, 5]. Regardless, robotic surgery is now an established clinical option, with multiple companies investing in these technologies [6].

One of the main challenges for robotic surgery in clinical practice is training of surgeons [7]. In [8] Herron et al. laid out guidelines for training and clinical practice of robotic surgery, highlighting the importance of a structured training curriculum, expert guidance and hands-on experience. Later in [9] Lee et al. emphasized the importance of skill-based assessments, to be prioritized for credentialing over the number of completed surgeries. Stegemann at al. in [10] laid the foundations for a unified approach to robotic training in the US, which was later validated and proposed as a standard training curriculum in [11]. Similar efforts were undertaken in Europe by Volpe et al. [12], and [13] reports on yet another attempt to defined a unified curriculum from an international panel. Despite these efforts, there is still no unified training system across disciplines and schools [14, 15].

In more recent years there has been an increased interest in the use of virtual reality (VR) simulators [16–18]. Using VR instead of a physical system comes with intrinsic advantages, namely by allowing complete control of all the parameters of a simulation, which in turn provides a more consistent mode of skill acquisition and mastery [19] and enables direct measurements of metrics of performance within the VR environment [20]. However, there also is the added challenge that comes from not necessarily being able to accurately simulate what would happen in a real world scenario, and simulation fidelity as well as its effectiveness in increasing skill in surgeons always need to be evaluated [21].

Several papers, including a few reviews, have been published on the topic of using VR for training in robotic surgery, with mixed findings. Lallas et al. evaluated results from the state of the art for four VR simulators, and highlighted their advantages over box trainers, which are low fidelity models and do not provide feedback on performance to the trainee, and animal models, which are closer to a real surgery scenario but are often simply not available [19]. VR systems on the other hand would be easily available and provide standardized metric and access to anatomic variants for different operations. A survey from practitioners of the field however also highlighted resistance to the use of VR systems because of their cost [19].

Hung et al. compared inanimate, virtual reality and in vivo methods for training [22]. They evaluated performance differences between novices and experts and measured the association of performance across the three methods. All methods showed a difference between experts and novices, and performance in the inanimate task correlated strongly with performance in the other methods.

More recently Bric et al. reviewed literature on the use of Virtual Reality for training on the Da Vinci robot [23]. They reported that some commercially available VR simulators are able to assess robotic training skill and can improve skill training at a level comparable to dry lab training. However, in the same year another review from Moglia et al. highlighted limitations in VR training, particularly in the fact that there is still not reported evidence of skills transfer from simulation to clinical surgery on real patients [24]. Authors also pointed out how evaluating trainers based on an overall score only can yield mixed results when tasks of different levels of difficulty are considered. This was more recently also asserted by Brook et al. in their comprehensive review on training for robotic surgery in urology [15], and by a small number of studies that found that VR training does not necessarily transfer to good skills on the robot for certain tasks (e.g., [25]).

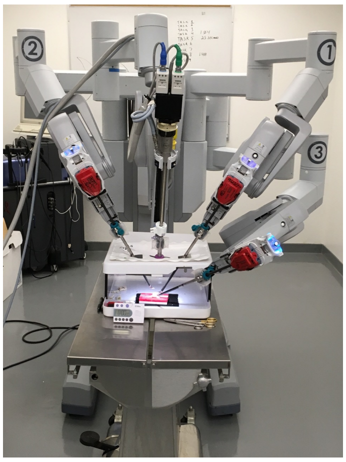


Fig. 1. da Vinci S surgical robot used for training with the UTSW proficiency based curriculum.

In summary, previous work has highlighted potential for VR training as a good tool for assessment, but evaluation of its effectiveness is still somewhat limited and often too generalized. In this paper, we aim to address this issue by evaluating the effect of one of the most popular VR training system, the da Vinci Skills Simulator, on performance with a robotic trainer in a standardized curriculum at the University of Texas Southwestern Medical Center. This effort was also motivated by recent results [26] which suggested a (statistically) significant difference in muscle activation and economy of volume between virtual reality simulations and dry lab training.

## 2. Methodology

This study was conducted at the University of Texas Southwestern Medical Center at Dallas in the Southwestern Center for Minimally Invasive Surgery's robotic surgical skills training laboratory, using a da Vinci robot (standard system; Intuitive Surgical, Inc., Sunnyvale, CA; Fig. 1). The da Vinci robot consists of a master-side manipulator with stereo view and a patient-side robot which could be used to manipulate the physical training environment.

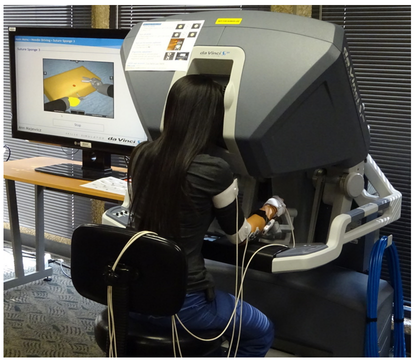
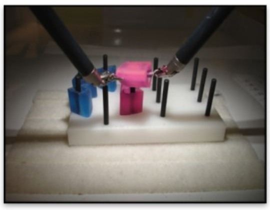


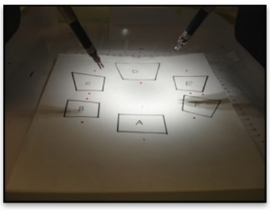
Fig. 2. VR training system for robotic skills consisting of the da Vinci Si master console and simulated exercises developed by Mimic Technologies, Inc.

### 2.1. *Established Robotics Curriculum*

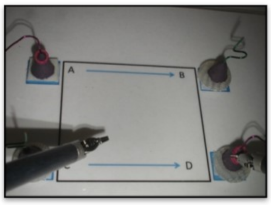
The University of Texas Southwestern has established a proficiency based curriculum for robotic surgery training [27–30], which we will refer to as the UTSW curriculum, and has been part of the teaching practice since 2012. It is based on a set of unique robotic skills that were identified by a panel of experts. The training protocol includes an online tutorial, one half-day hands on session, and nine inanimate training exercises. Trainees are scored for each task based on the time to completion and errors, with a normalized score being defined as the task score divided by



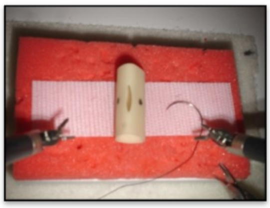
(a) Task 1: Peg Transfer.



(b) Task 2: Clutch/Camera Movement.

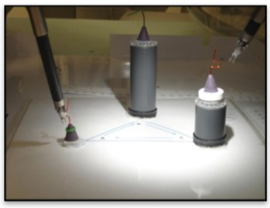


(c) Task 3: Rubber Band Transfer.



(d) Task 4: Suture (Simple).

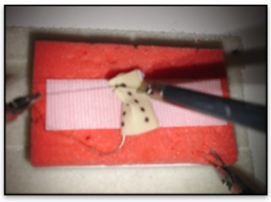
(e) Task 5: Clutch/Camera Peg Transfer.



(f) Task 6: Stair Rubber Band Transfer.

(g) Task 7: Running/Cutting Rubber Band.

(h) Task 8: Pattern Cut.



(i) Task 9: Suture (Running).

Fig. 3. Physical Training Tasks in the UTSW Proficiency-based Surgical Robotics Curriculum.

a proficiency score as obtained by an expert, and a composite score being defined as the sum of normalized scores for each task. Pre-test scores are recorded on a single proctored repetition of each task at the end of the hands on session, while post-test scores are recorded after training.

Table 1. Proficiency scores and times for the UTSW curriculum.

| Task | Score | Time (s) |
|---|---|---|
| 1: Peg Transfer | 97 | 66 |
| 2: Clutch/Camera Movement | 104 | 67 |
| 3: Rubber Band Transfer | 94 | 68 |
| 4: Suture (Simple) | 100 | 69 |
| 5: Clutch/Camera Peg Transfer | 106 | 70 |
| 6: Stair Rubber Band Transfer | 104 | 71 |
| 7: Running/Cutting Rubber Band | 110 | 72 |
| 8: Pattern Cut | 101 | 73 |
| 9: Suture (Running) | 133 | 74 |

The experiments presented in this paper were based on this curriculum. Figure 3 shows the nine inanimate exercises that are part of the protocol used for training. Tasks are ordered by level of complexity, with the last task being the most difficult. In detail:

- *Task 1 - Peg Transfer:* transfer six rubber pegs from the left side of a pegboard to the right side, and then transfer them back. The camera is static in this task. Pegs must be transferred from one needle driver to the other when switching sides;
- *Task 2 - Clutch and Camera Movement:* move and navigate the camera to focus it on six geometrical shapes, zooming on each enough to cut off the red dots placed above and below each shape;
- *Task 3 - Rubber Band Transfer:* transfer a rubber band through custom designed curved wire posts, without deforming the wire;
- *Task 4 - Suture (Simple):* perform a simple suture on Penrose drain;
- *Task 5 - Clutch and Camera Peg Transfer:* transfer pegs through metal posts on a custom-designed wooden board, clutching and moving the camera.

- *Task 6 - Stair Rubber Band Transfer:* transfer a rubber band through custom designed curved wire posts placed at different heights, clutching and moving the camera as needed;
- *Task 7 - Running and Cutting Rubber Band:* cut a 12 cm long rubber band model at 10 marked spots, clutching as needed, and dropping and re-grasping the rubber band between cuts;
- *Task 8 - Pattern Cut:* cut a circular pattern on a testing gauze, moving the camera and clutching as needed;
- *Task 9 - Suture (Running):* perform a running suture on a Penrose drain.

This curriculum has been used at UT Southwestern since 2012 and has trained 365 surgical residents.

### 2.2. *Virtual Reality Pre-Training Curriculum*

A total of 9 tasks were selected from the da Vinci Skills Simulator (dVSS, Intuitive Surgical, Sunnyvale, CA), where were developed with Mimic Technologies Inc. (Seattle, WA). The tasks were chosen as those that most closely represent skills needed in the UTSW curriculum, and were selected from sets previously validated for face and construct validity in [20] and [31] (these papers did not however investigate the effect of the dVSS on the development of surgical skills). Target proficiency scores and completion time were defined for each task [20] and are reported here for the selected tasks (Table 2).

Table 2. Proficiency scores and times for the selected tasks from the da Vinci Skills Simulator [20].

| Task | Score | Time (s) |
|---|---|---|
| 1: Pick and Place | 95 | 28 |
| 2: Camera Targeting 2 | 87 | 87 |
| 3: Peg Board 1 | 87 | 63 |
| 4: Tubes | 71 | 207 |
| 5: Match Board 3 | 56 | 290 |
| 6: Ring & Rail 2 | 75 | 185 |
| 7: Ring Walk 3 | 63 | 142 |
| 8: Energy Switching | 83 | 81 |
| 9: Suture Sponge | 68 | 285 |

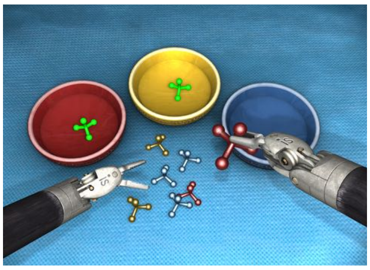
Figure 4 shows the VR tasks that were included, in particular:

- *Task 1 - Pick and Place:* place objects inside containers of the corresponding color (comparable with physical task 1);
- *Task 2 - Camera Targeting 2:* move and navigate the camera while working in a large volume and manipulating objects (comparable with physical tasks 2 and 5);
- *Task 3 - Peg Board 1 & 2:* pick up rings from pegs, performs a transfer from instrument to another, and place the rings on a new peg (comparable with physical task 1);
- *Task 4 - Tubes:* drive needle through fixed targets on cylindrical deformable structure, to simulate tissue manipulation (comparable with physical task 4);
- *Task 5 - Match board 3:* place objects at specific sites on a match board with removable covers (comparable with physical task 5);
- *Task 6 - Ring and Rail 2:* move a ring across a twisted rail to test alternating camera and instrument control (comparable with physical tasks 2, 5 and 6);
- *Task 7 - Ring Walk 3:* similar to the previous task, but in an occluded environment, leading to a more challenging camera control as well as the need to remove obstacles (comparable with physical tasks 2, 5 and 6);
- *Task 8 - Energy Switching:* switch between monopolar and binopolar energy while working on a dissection task (comparable with physical task 8).
- *Task 9 - Suture Sponge:* perform a running suture on a virtual sponge (comparable with physical task 9).

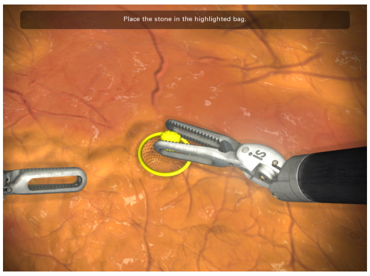
It can be noted that there is not necessarily a one on one correspondence between the physical and VR tasks. This is caused by the fact that we were constrained to only choose from the VR training tasks that were already present in the commercial da Vinci Skills Simulator, rather than develop new ones. However, as shown in the detailed explanation of both the VR and physical tasks, the overall set of skills being targeted was the same, consistently with the skill-based approach that was used in the development of the dVSS [17].

### 2.3. *Experimental Protocol*

A total of 107 third year residents (40 male, 67 female) from general surgery (29) , urology (12) and gynecology (66) were recruited to complete surgical robotics training at UTSW between 2015 and 2018 . As this study originated as part of an educational improvement effort, participants were divided, on a voluntary basis, into two groups: one first went through the VR training program (VR/y), and then took part in the UTSW curriculum, while the other group only took part in the UTSW curriculum (VR/n). We elected to make the participation voluntary since the UTSW curriculum is part of the formative experience offered to residents at the University of Texas Southwestern, and we did not know if the additional VR training would



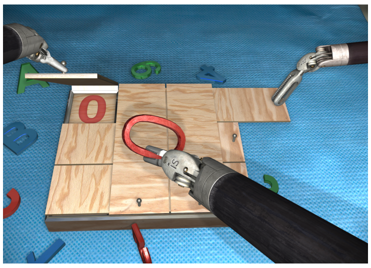
(a) Virtual Task 1: Pick and Place.



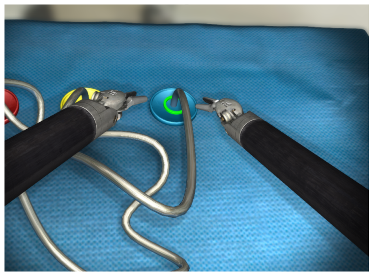

(b) Virtual Task 2: Camera Targeting 2.

(c) Virtual Task 3: Peg Board 1 & 2.

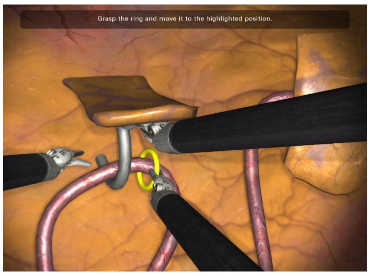

(d) Virtual Task 4: Tubes.



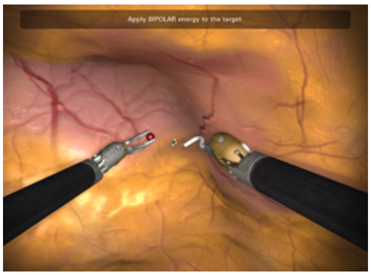
(e) Virtual Task 5: Matchboard 3.



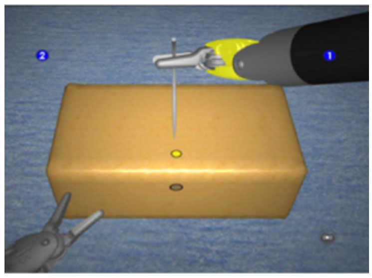
(f) Virtual Task 6: Ring and Rail 2.




(g) Virtual Task 7: Ring Walk 3.




(h) Virtual Task 8: Energy Switching.



(i) Virtual Task 9: Suture Sponge.

Fig. 4. Virtual Reality Tasks used during Pre-Training. A total of 9 tasks were used for pre-training on the da Vinci Skills Simulator (Si console), using virtual reality tasks developed by Mimic Technologies, Inc. Images obtained with permission.

be beneficial, have no effect or perhaps even be harmful to their training.

A total of 18 participants (10 from general surgery, 5 from urology and 3 from gynecology; 12 males, 6 females) elected to take part in the VR training, leaving 89 for the second group. Performance was evaluated based on the global and task scores during robotic task performance, as defined by the UTSW curriculum, as well as the number of repetitions to proficiency. Pre-test and Post-test scores (taken before and after participants took part in the UTSW curriculum) were considered for comparison. The data analysis of unidentified data for this publication occurred after obtaining approval from the UTSW IRB as an exempt study. Scores from the virtual reality simulator were not recorded and therefore not available for analysis; however, all subjects were required to meet or exceed a proficiency score for each of the VR tasks, as reported in Section 2.2. Both groups were evaluated using the same scoring method defined by the UTSW curriculum.

### 2.4. *Statistical Data Analysis*

A comparison between the two groups was performed in terms of both the global score and individual task scores from the physical training with the robot according to the UTSW curriculum. Due to the fact that participation in the extra VR training was voluntary, the overwhelming majority of participants fell in the VR/n condition, leading to an unbalanced design. Two different analyses were performed on the task scores:

- **Global score analysis:** in this case the VR training represents one between-subject factor, while the two measurements of score before and after the physical training on the robot (Pre/Post) represent a within-subject factor. This led us to compare the

means of the global score with a two-way mixed model ANOVA;
- **Task normalized scores analysis:** in addition to the previous two factors, in this case the different tasks represent an additional within-subject factor, leading to a three-way mixed model ANOVA with one between-subject (VR) and two within-subject (Pre/Post, task) factors.

Unbalanced ANOVA designs come with several complications [32], among which are a loss of robustness to violations of assumptions on normality and heteroskedasticity. Furthermore, a power analysis performed in G* Power [33] showed that for both designs the total number of subjects needed to have 0.8 power is 20 and 27, respectively, assuming an effect size (partial $\eta^2 = 0.1$) and a significance level $\alpha = 0.05$.

For these reasons we decided to subsample the VR/n condition, and analyse our data through balanced ANOVAs, which are robust to violations of normality [34,35] and homogeneity of variance [36]. Of the 89 participants falling under the VR/n condition, 18 were selected, matching the 18 in the VR/y condition in gender and specialization (10 from general surgery, 5 from urology and 3 from gynecology; 12 males, 6 females), for a total of 36 participants. When more than one trainee in the VR/n condition matched a certain VR/y trainee, one was randomly selected. A two-way mixed ANOVA was then performed for the global score, and a three-way mixed ANOVA for the normalized local score. Huynh-Feldt correction of p-values was applied when non sphericity was present.

The medians of the number of repetitions during training were compared through a Mann-Whitney U test. All analysis was done in R.

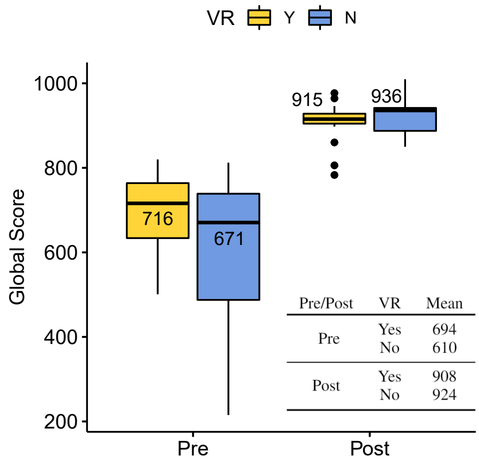



Fig. 5. Global normalized scores before (Pre) and after (Post) administration of the UTSW curriculum (calculated as sum of the normalized scores for each task).

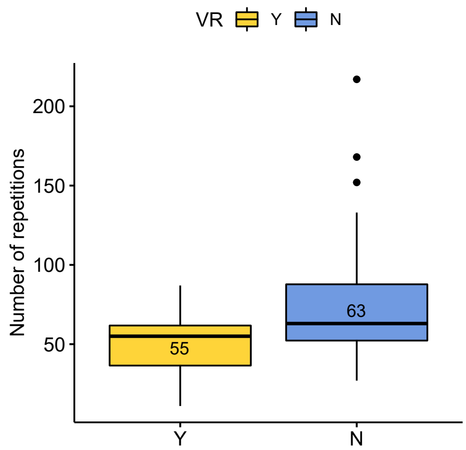



Fig. 6. Number of repetitions to proficiency over all tasks.

## 3. Results and Discussion

Figure 5 shows bar plots for the global normalized scores for pre-test and post-test for both groups. Visual inspection suggests a higher pre-test score for the VR condition when compared to the no VR condition, while no evident difference between the two emerges for the post-test. There also seems to be a noticeable difference between pre and post-test for both conditions. Figure 6 shows a bar plot for the number of repetitions during the pre-test. The plot shows a tendency for a lower number of repetitions on average for the VR condition.

An overview of task-wise scores can be seen in Fig. 7. VR training seems to yield slightly better scores for most tasks during the pre-test, with differences being less evident as task difficulty increases (e.g., task 8 has similar performance for both conditions). No clear difference between the two conditions can be observed for the post-test score. When it comes to comparing pre and post-test for each VR condition, there seems to be an increase in performance for both, although it is not as marked for the VR condition.

Table 3 shows the statistical results of the two way mixed ANOVA for global scores. There is a significant interaction effect ($p = 0.039$) as well as a strongly significant main effect for the Pre/Post (i.e., before or after physical training with the robot) factor ($p = 4.2e - 13$), while no evidence of main effect appears for the VR factor. Notably, the effect size, estimated through the generalized eta squared ${\eta^2}_G$ [37], is much smaller for the interaction than for the physical training main effect. Because an interaction is present, we decided to look at the simple main effects by running independent t-tests for the VR factor at each level of the Pre/Post factor, with Holm-Bonferroni correction. Results yielded non-significant p-values for the effect of VR in both cases ($p = 0.166$ for Pre and $p = 0.288$ for Post).

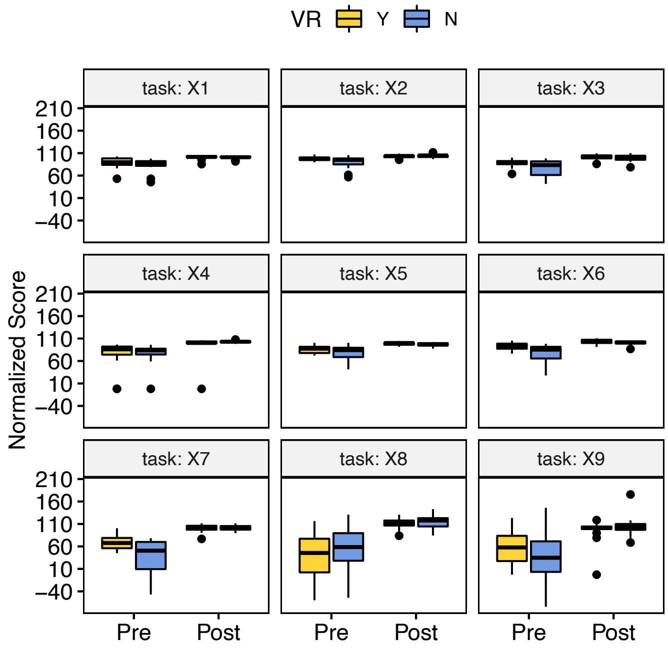



Fig. 7. Normalized scores before (Pre) and after (Post) administration of the UTSW curriculum, for each task, for both the VR and no VR conditions.

Table 3. Results for two-way mixed ANOVA on global scores. Significant p-values in bold.

| Effect | $DF_n$ | $DF_d$ | $F$ | $p$ | ${\eta^2}_G$ |
|---|---|---|---|---|---|
| VR | 1 | 34 | 1.716 | 0.199 | 0.027 |
| Pre/post | 1 | 34 | 128.677 | **4.2e − 13** | 0.628 |
| VR:Pre/post | 1 | 34 | 4.618 | **0.039** | 0.057 |

Table 4. Results for three-way mixed ANOVA on normalized task scores. Significant p-values in bold.

| Effect | $DF_n$ | $DF_d$ | $F$ | $p$ | ${\eta^2}_G$ |
|---|---|---|---|---|---|
| VR | 1 | 34 | 1.716 | 0.199 | 0.008 |
| Pre/post | 1 | 34 | 128.677 | **4.2e − 13** | 0.324 |
| task | 3.77 | 128.11 | 12.549 | **1.20e − 07** | 0.125 |
| VR:Pre/post | 1 | 34 | 4.618 | **0.039** | 0.017 |
| VR:task | 3.77 | 128.11 | 1.998 | 0.103 | 0.022 |
| Pre/Post:task | 3.21 | 109.07 | 22.354 | **7.87e − 12** | 0.177 |
| VR:Pre/Post:task | 3.21 | 109.07 | 1.837 | 0.141 | 0.017 |

Table 5. Results for post hoc two-way ANOVAs on normalized task scores. Significant p-values in bold.

| Pre/Post | Effect | $DF_n$ | $DF_d$ | $F$ | $p$ (Holm) | ${\eta^2}_G$ |
|---|---|---|---|---|---|---|
| Pre | VR | 1 | 34 | 3.19 | 0.332 | 0.028 |
| Pre | task | 3.42 | 116.14 | 19.2 | **9.71e − 11** | 0.283 |
| Pre | VR:task | 3.42 | 116.14 | 2.03 | 0.211 | 0.04 |
| Post | VR | 1 | 34 | 1.17 | 0.496 | 0.006 |
| Post | task | 3.41 | 115.87 | 5.68 | **0.000683** | 0.12 |
| Post | VR:task | 3.41 | 115.87 | 1.40 | 0.244 | 0.033 |

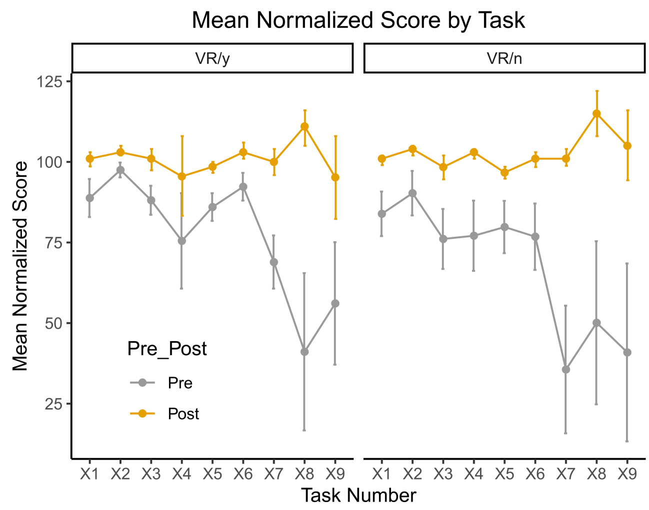



Fig. 8. Mean normalized scores by task, before (Pre) and after (Post) administration of the UTSW curriculum, for both the VR and no VR conditions

Results for the three-way ANOVA on the task-normalized scores are shown in Table 4. Non-sphericity was determined for the factor task, which was corrected with Huynh-Feldt. The three-way interaction was non significant, and two-way interactions were significant between

VR and Pre/Post ($p = 0.039$) and Pre/Post and task ($p = 6.32e - 11$). We also observed significant main effects for Pre/Post ($p = 4.2e - 13$) and task ($p = 1.20e - 07$) factors. In order to look further into the significant two-way interaction effects, we did two-way ANOVAs at each level of the Pre/Post factor, and considered the main effects of VR and task. Factor VR was non significant at each level of Pre/Post ($p = 0.332$ for Pre and $p = 0.496$ for Post), while factor task was significant for both ($p = 2.24e - 10, {\eta^2}_G = 0.283$ and $p = 0.005, {\eta^2}_G = 0.12$ respectively).

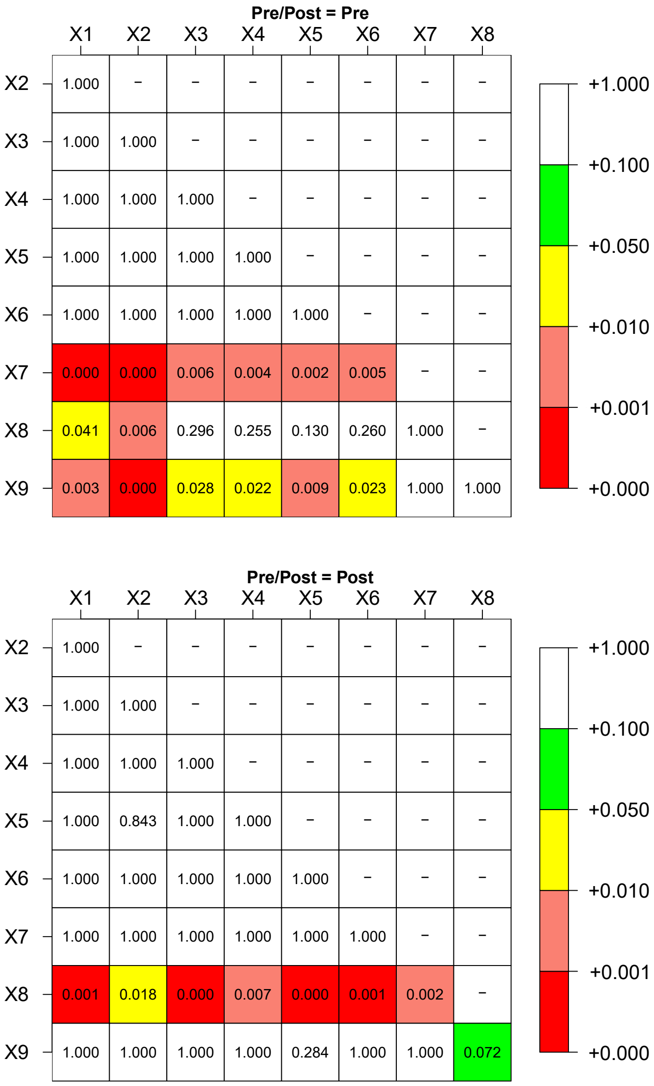



Fig. 9. Post-hoc pairwise comparison: p-values (Holm adjusted).

To better evaluate the effect of the task factor on scores, we focused on the VR/n condition and considered the Pre and Post conditions of the Pre/Post factor separately. Figure 8 shows an overview of the normalized score for each task under both conditions. Tasks 1 to 6 appear to obtain similar scores for the Pre condition, with tasks 1 and 2 having slightly higher means, while tasks 7 to 9 yield noticeably lower mean scores. Normalized scores for the Post condition are relatively homogeneous, with the exception of task 8 which has a noticeably higher mean. Figure 9 shows the outcome of the pairwise paired t-tests, with p values adjusted using the Holm-Bonferroni correction.

Finally, a Mann-Whitney U test on the number of repetitions to proficiency showed a weakly significant $p$ value ($p = 0.09$, effect size $r = 0.285$) with a median of 63 for the VR/n condition and 55 for the VR/y condition.

To summarize, our statistical analysis found no significant effect of VR training on scores, neither globally nor at the task level. We did observe a significant improvement of score with the physical training, which confirms its validity consistently with results reported in [28–30]. Finally, we observed a significant effect of the task factor, especially for the Pre session, with tasks 7 to 9 proving to be more difficult, although task 8 actually yielded a significantly higher score in the Post session. VR training did show some effect on the number of repetitions to proficiency during the physical training, with significance at the 0.1 level. This means that while VR did not show indication of improving performance, it might be able to speed up physical training.

## 4. Conclusions

In this paper we presented a study on the effect of virtual reality (VR) pre-training with the da Vinci Skills Simulator (dVSS) on the measured outcome of the robotic training curriculum developed at the Univeristy of Texas Southwestern (UTSW). This curriculum evaluates trainee performance before and after training on a physical robot through a standardized score system. One group of participants underwent additional training with the dVSS, before they were administered the UTSW curriculum, while the other group was only administered the UTSW curriculum. Scores from the UTSW curriculum were used to evaluated performance in both groups. We did not observe any significant effect of training with the dVSS on performance, while there were some indications of a reduction in the number of repetitions to proficiency during training (weakly significant at $p < 0.1$).

Analysis of individual task scores showed that tasks 7 (running/cutting rubber band) and 9 (running suture) yielded significantly lower scores when compared to tasks 1 to 6 (peg transwer, clutch/camera movement, rubber band transfer, simple suture, clutch/camera peg transfer and stair rubbed band transfer) before the physical training occurred, while task 8 (pattern cut) yielded a significant lower score when compared to tasks 1 and 2, suggesting a higher level of difficulty for these tasks. Interestingly, no significant difference was observed on tasks scores after the physical training occurred with the exception of task 8, which yielded a significantly higher score than all other tasks.

These results suggest that the da Vinci Skills Simulator is not yet able to completely replace the training offered by the UTSW curriculum , but shows potential for reducing training time on the physical simulator. While this by itself could beneficial, buying a VR training system for robotic surgery comes with a significant additional cost of at least $80000 [38], much higher than the costs associated with the physical training curriculum for which material cost is $2227 [29]. This might make buyers question if a potential mild reduction in training times justifies the additional investment required. Additionally, it is interesting to note that of the 107 residents involved in this study, only 18 elected to go through the additional VR pre-training when given the choice, indicating an opportunity to better incentive VR training, perhaps through competition or a more game-like interface [39]. Ultimately, we believe that the outcome of this work indicates a possible need to enhance the da Vinci Skills Simulator VR training to either more closely reflect physical reality to deliver its full potential in terms of training benefit, or hone very specific skills in a more abstract setting. It also suggests the benefit of a thorough proficiency-based evaluation of VR training systems for surgical robotics.

## Acknowledgment

We would like to acknowledge the dVSS simulator loan program at Intuitive Surgical which enabled this study, as well as Mimic Technologies for providing images for this publication.

## References

[1] A. Ng and P. Tam, Current status of robot-assisted surgery, *Hong Kong Med J* **20**(3) (2014) 241–250.
[2] K. H. Sheetz, J. Claflin and J. B. Dimick, Trends in the adoption of robotic surgery for common surgical procedures, *JAMA network open* **3**(1) (2020) e1918911–e1918911.
[3] J. J. Leow, S. L. Chang, C. P. Meyer, Y. Wang, J. Hanske, J. D. Sammon, A. P. Cole, M. A. Preston, P. Dasgupta, M. Menon *et al.*, Robot-assisted versus open radical prostatectomy: a contemporary analysis of an all-payer discharge database, *European urology* **70**(5) (2016) 837–845.
[4] J. W. Yaxley, G. D. Coughlin, S. K. Chambers, S. Occhipinti, H. Samaratunga, L. Zajdlewicz, N. Dunglison, R. Carter, S. Williams, D. J. Payton *et al.*, Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study, *The Lancet* **388**(10049) (2016) 1057–1066.
[5] D. Jayne, A. Pigazzi, H. Marshall, J. Croft, N. Corrigan, J. Copeland, P. Quirke, N. West, T. Rautio, N. Thomassen *et al.*, Effect of robotic-assisted vs conventional laparoscopic surgery on risk of conversion to open laparotomy among patients undergoing resection for rectal cancer: the rolarr randomized clinical trial, *Jama* **318**(16) (2017) 1569–1580.
[6] R. Perez and S. Schwaitzberg, Robotic surgery: finding value in 2019 and beyond, *Ann Laparosc Endosc Surg* **4** (2019) p. 51.
[7] S. D. Herrell and J. A. Smith JR, Robotic-assisted laparoscopic prostatectomy: what is the learning curve?, *Urology* **66**(5) (2005) 105–107.
[8] D. Herron, M. Marohn *et al.*, A consensus document on robotic surgery, *Surgical endoscopy* **22**(2) (2008) 313–325.
[9] J. Y. Lee, P. Mucksavage, C. P. Sundaram and E. M. McDougall, Best practices for robotic surgery training and credentialing, *The Journal of urology* **185**(4) (2011) 1191–1197.
[10] A. P. Stegemann, K. Ahmed, J. R. Syed, S. Rehman, K. Ghani, R. Autorino, M. Sharif, A. Rao, Y. Shi, G. E. Wilding *et al.*, Fundamental skills of robotic surgery: a multi-institutional randomized controlled trial for validation of a simulation-based curriculum, *Urology* **81**(4) (2013) 767–774.
[11] R. M. Satava, D. Stefanidis, J. S. Levy, R. Smith, J. R. Martin, S. Monfared, L. R. Timsina, A. W. Darzi, A. Moglia, T. C. Brand *et al.*, Proving the effectiveness of the fundamentals of robotic surgery (frs) skills curriculum: a single-blinded, multispecialty, multi-institutional randomized control trial, *Annals of surgery* **272**(2) (2020) 384–392.
[12] A. Volpe, K. Ahmed, P. Dasgupta, V. Ficarra, G. Novara, H. van der Poel and A. Mottrie, Pilot validation study of the european association of urology robotic training curriculum, *European urology* **68**(2) (2015) 292–299.
[13] K. Ahmed, R. Khan, A. Mottrie, C. Lovegrove, R. Abaza, R. Ahlawat, T. Ahlering, G. Ahlgren, W. Artibani, E. Barret *et al.*, Development of a standardised training curriculum for robotic surgery: a consensus statement from an international multidisciplinary group of experts, *BJU international* **116**(1) (2015) 93–101.
[14] A. N. Sridhar, T. P. Briggs, J. D. Kelly and S. Nathan, Training in robotic surgery—an overview, *Current urology reports* **18**(8) (2017) 1–8.
[15] N. R. Brook, P. Dell'Oglio, R. Barod, J. Collins and A. Mottrie, Comprehensive training in robotic surgery, *Current opinion in urology* **29**(1) (2019) 1–9.
[16] J. M. Albani and D. I. Lee, Virtual reality-assisted robotic surgery simulation, *Journal of endourology* **21**(3) (2007) 285–287.
[17] A. S. Sethi, W. J. Peine, Y. Mohammadi and C. P. Sundaram, Validation of a novel virtual reality robotic simulator, *Journal of endourology* **23**(3) (2009) 503–508.
[18] C. Perrenot, M. Perez, N. Tran, J.-P. Jehl, J. Felblinger, L. Bresler and J. Hubert, The virtual reality

simulator dv-trainer® is a valid assessment tool for robotic surgical skills, *Surgical endoscopy* **26**(9) (2012) 2587–2593.

[19] C. D. Lallas, Davis and J. W. Members of the Society of Urologic Robotic Surgeons, Robotic surgery training with commercially available simulation systems in 2011: a current review and practice pattern survey from the society of urologic robotic surgeons, *Journal of endourology* **26**(3) (2012) 283–293.

[20] C. Lyons, D. Goldfarb, S. L. Jones, N. Badhiwala, B. Miles, R. Link and B. J. Dunkin, Which skills really matter? proving face, content, and construct validity for a commercial robotic simulator, *Surgical endoscopy* **27**(6) (2013) 2020–2030.

[21] P. H. Cosman, P. C. Cregan, C. J. Martin and J. A. Cartmill, Virtual reality simulators: current status in acquisition and assessment of surgical skills, *ANZ journal of surgery* **72**(1) (2002) 30–34.

[22] A. J. Hung, I. S. Jayaratna, K. Teruya, M. M. Desai, I. S. Gill and A. C. Goh, Comparative assessment of three standardized robotic surgery training methods, *BJU international* **112**(6) (2013) 864–871.

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

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

[25] V. Phé, S. Cattarino, J. Parra, M.-O. Bitker, V. Ambrogi, C. Vaessen and M. Rouprêt, Outcomes of a virtual-reality simulator-training programme on basic surgical skills in robot-assisted laparoscopic surgery, *The International Journal of Medical Robotics and Computer Assisted Surgery* **13**(2) (2017) p. e1740.

[26] Z. Wang, M. Kasman, M. Martinez, R. Rege, H. Zeh, D. Scott and A. M. Fey, A comparative human-centric analysis of virtual reality and dry lab training tasks on the da vinci surgical platform, *Journal of Medical Robotics Research* **4**(03n04) (2019) p. 1942007.

[27] G. Dulan, R. V. Rege, D. C. Hogg, K. K. Gilberg-Fisher, S. T. Tesfay and D. J. Scott, Content and face validity of a comprehensive robotic skills training program for general surgery, urology, and gynecology, *The American journal of surgery* **203**(4) (2012) 535–539.

[28] G. Dulan, R. V. Rege, D. C. Hogg, K. M. Gilberg-Fisher, N. A. Arain, S. T. Tesfay and D. J. Scott, Proficiency-based training for robotic surgery: construct validity, workload, and expert levels for nine inanimate exercises, *Surgical endoscopy* **26**(6) (2012) 1516–1521.

[29] G. Dulan, R. V. Rege, D. C. Hogg, K. M. Gilberg-Fisher, N. A. Arain, S. T. Tesfay and D. J. Scott, Developing a comprehensive, proficiency-based training program for robotic surgery, *Surgery* **152**(3) (2012) 477–488.

[30] N. A. Arain, G. Dulan, D. C. Hogg, R. V. Rege, C. E. Powers, S. T. Tesfay, L. S. Hynan and D. J. Scott, Comprehensive proficiency-based inanimate training for robotic surgery: reliability, feasibility, and educational benefit, *Surgical endoscopy* **26**(10) (2012) 2740–2745.

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

[32] R. G. Shaw and T. Mitchell-Olds, Anova for unbalanced data: an overview, *Ecology* **74**(6) (1993) 1638–1645.

[33] F. Faul, E. Erdfelder, A.-G. Lang and A. Buchner, G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, *Behavior research methods* **39**(2) (2007) 175–191.

[34] E. Schmider, M. Ziegler, E. Danay, L. Beyer and M. Bühner, Is it really robust?, *Methodology* (2010).

[35] M. J. Blanca, R. Alarcón, J. Arnau, R. Bono and R. Bendayan, Non-normal data: Is anova still a valid option?, *Psicothema* **29**(4) (2017) 552–557.

[36] R. Refinetti, Demonstrating the consequences of violations of assumptions in between-subjects analysis of variance, *Teaching of Psychology* **23**(1) (1996) 51–54.

[37] R. Bakeman, Recommended effect size statistics for repeated measures designs, *Behavior research methods* **37**(3) (2005) 379–384.

[38] A. M. Hertz, E. I. George, C. M. Vaccaro and T. C. Brand, Head-to-head comparison of three virtual-reality robotic surgery simulators, *JSLS: Journal of the Society of Laparoendoscopic Surgeons* **22**(1) (2018).

[39] D. A. Hashimoto, E. D. Gomez, L. Beyer-Berjot, A. Khajuria, N. N. Williams, A. Darzi and R. Aggarwal, A randomized controlled trial to assess the effects of competition on the development of laparoscopic surgical skills, *Journal of surgical education* **72**(6) (2015) 1077–1084.