

**Are basic robotic surgical skills transferable from the simulator to the operating room?
A randomized, prospective, educational study**

Ahmad Almarzouq^{1*}; Jason Hu^{1*}; Yasser A. Nouredin^{1,2}; Anne Yin¹; Maurice Anidjar¹; Franck Bladou¹; Simon Tanguay¹; Wassim Kassouf¹; Armen G. Aprikian¹; Sero Andonian^{1,3}

¹Division of Urology, McGill University Health Centre, Montreal, QC, Canada; ²Department of Urology, Faculty of Medicine, Benha University, Benha, Egypt; ³Institute for Health Sciences Education, McGill University, Montreal, QC, Canada

**Equal contributors*

Acknowledgements: Grant received from The Kuwaiti Foundation for the Advancement of Science (KFAS).

Cite as: *Can Urol Assoc J* 2020 June 16; Epub ahead of print.
<http://dx.doi.org/10.5489/cuaj.6460>

Published online June 16, 2020

Abstract

Introduction: We aimed to assess the transferability of basic robotic skills from the simulator to the operating room while performing robotic-assisted radical prostatectomy (RARP prostatectomy).

Methods: Fourteen urology residents were randomized to two groups: group A was required to practice three sessions (nine tasks each) on the simulator, whereas group B was required to practice (same nine tasks) until they reached competency. Both groups were recorded while practicing on the da Vinci Surgical Skills Simulator (da Vinci Simulator). Both groups were then recorded while performing bladder mobilization during RARP. Senior residents from both groups were also recorded while performing urethro-vesical anastomosis during RARP. Recordings were assessed blindly using the validated GEARS tool by C-SATS. Spearman's correlation coefficient (ρ) was used to assess correlation between GEARS scores from practice sessions on the da Vinci Simulator and the GEARS scores from bladder mobilization and urethro-vesical anastomosis during RARP.

Results: There was no difference in total GEARS scores between the two groups in the operating room. Total GEARS scores for “ring and rail 2” and “suture sponge” tasks correlated with the total GEARS scores during urethro-vesical anastomosis ($\rho=0.86$, $p=0.007$; $\rho=0.90$, $p=0.002$, respectively). GEARS' efficiency component during “energy and dissection” task on the da Vinci Simulator correlated with GEARS' efficiency component during bladder

mobilization ($\rho=0.62$, $p=0.03$). GEARS' force sensitivity component during "ring and rail 2" and "dots and needles" tasks on the da Vinci Simulator correlated with GEARS' force sensitivity component during bladder mobilization ($\rho=0.58$, $p=0.047$; $\rho=0.65$, $p=0.02$, respectively).

Conclusions: Objective assessments of urology residents on the da Vinci Surgical Skills Simulator tasks ring and rail 2 and suture sponge correlated with their objective assessments of bladder mobilization and urethro-vesical anastomosis. Therefore, basic robotic skills could be transferred from the simulator to the operating room.

Introduction

Prostate cancer is the most common malignancy in men and radical prostatectomy represents a standard treatment for clinically localized prostate cancer with acceptable complication rates and functional outcomes [1,2]. Since the introduction of robotic surgery in the early 2000s, it has been widely adopted in urology. Recently, Robotic-assisted radical prostatectomy (Robotic Prostatectomy) has become a standard procedure for localized prostate cancer. However, similar to the open approach, it has a steep learning curve. Although there is no widely accepted definition of a learning curve for Robotic Prostatectomy, several authors attempted to define it. Several outcomes could be used to define the number of cases required to achieve the plateau on the learning curve. Estimated blood loss was found to plateau in novice surgeons at 100-200 cases, operative time plateaued between 40-120 cases, length of stay plateaued at 200 cases, and positive surgical margins plateaued between 50-1600 cases [3-6]. Despite these findings being established at an early stage of Robotic Prostatectomy, the fact remains that Robotic Prostatectomy is not an easy procedure to learn and poses a challenge to surgical trainees and trainers. This challenge is further compounded by the reduction in working hours and the recent introduction of competency-based learning in urology training programs. Therefore, competency threshold in basic robotic skills is needed to assess competency of urology residents in basic robotic skills. To date, there is no validated curriculum that has been shown to determine competency of urology residents.

Multiple studies attempted to define some elements of the challenge. First, virtual reality (VR) simulators were shown to correlate well with dry lab performances and are on par as a training modality and are less costly [7,8]. Second, simulators including VR simulators have been shown to improve technical skills with more training regardless of the curriculum implemented [9-12]. Third, an objective form of assessment was established with the introduction and validation of the Global Evaluative Assessment of Robotic Skills (GEARS) scale and was shown to have good correlation with operative performance in surgeons and higher scores resulted in better surgical outcomes [13-15].

The assessment of these surgical performances using objective tools is very time consuming. In these studies, the volume of performances is sometimes too large to be analyzed in a timely manner by experts. Therefore, crowdsourced scoring of these videos is being utilized and has been validated to correlate well with expert scoring in Robotic Prostatectomy [16-18]. However, to date there are no studies demonstrating that basic robotic skills could be transferred directly from the simulator to the operating room.

Therefore, the first objective of this study was to establish competency cut-offs on the da Vinci Surgical Skills Simulator (da Vinci Simulator) for basic robotic skills. The second objective was to correlate Global scores from the da Vinci Simulator for each task with the GEARS evaluations of the same tasks. The third objective was to assess the transferability of basic robotic skills from the da Vinci Simulator to the operating room using the GEARS evaluations.

Methods

After obtaining Institutional Research Ethics Board approval, 14 urology residents and five attending urologists (experts) were enrolled in this longitudinal study that spanned from June 2015 till January 2019. The da Vinci Surgical Skills Simulator (da Vinci Simulator) (Intuitive Surgical, Inc., Sunnyvale, CA) was used to assess basic robotic surgical skills and to train robotic-naïve urology residents. The majority of residents at the time of enrollment had participated in bedside assistance for robotic procedures. However, none of them had previous exposure to the robotic console prior to the study. Based on previous construct validation studies [19], the following 9 tasks were used: Peg board level 2, match board level 2, needle targeting, ring and rail level 2, dots and needles level 1, suture sponge level 2, energy dissection level 1, ring walk level 3 and tubes. Two of the five experts had previously participated as experts in the previous validation study and had performed all 9 tasks [19]. At the start of present study, all 5 experts were recorded while performing the 9 tasks. Competency cut-off thresholds were established for the global scores for each of the 9 tasks on the da Vinci Simulator using the norm-referenced method (the mean of the expert group minus 1 standard deviation = competency cut-off) [20].

Participating residents were then randomized into two groups using (<https://www.randomizer.org/>). Group A was allowed to perform all 9 tasks 3 times on 3 separate days. Group B was allowed to perform all 9 tasks until competency was reached for each of the 9 tasks. Prior to starting their training on the da Vinci Simulator, all residents received the same 10-minute hands-on introduction session to become familiar with the clutches and finger controls of the simulator. In addition, prior to starting each task, residents watched a short tutorial video provided by the simulator explaining the objectives and the difficulty of each task. For each task, the simulator generated a total score ranging from 0 (worst) to 100 (best). The score is based on a set of pre-determined metrics including the following parameters: time to complete task,

economy of motion, instrument collision, excessive instrument force, master work space range, object drops, instruments out of view, missed targets, misplaced energy, broken vessels, and blood loss volume. In addition to the global scores, all of the tasks performed by both groups of residents and experts were recorded.

Once the simulator phase of the study was completed, robotic skills of the console naïve residents were assessed during Robotic Prostatectomy. For each resident, their first three bladder mobilizations during Robotic Prostatectomy were recorded. Once the residents became senior residents (Post-Graduate Years 4 and 5), their first three urethro-vesical anastomoses were also recorded. When the operating room phase concluded, all of the video recordings from the simulator and from the operating room were blinded and sent for objective assessment by crowd sourced group (C-SATs) using the validated Global Evaluative Assessment of Robotic Skills (GEARS) assessment tool with an overall score of 30. GEARS has 6 components including efficiency, bimanual dexterity, depth perception, force sensitivity, robotic control and autonomy; each with a maximum score of 5. The autonomy component was not assessed in this study since blinded endoscopic videos were used. Since the autonomy component was not used, the maximum overall GEARS score was 25.

There were 50 C-SATs reviewers, who blindly assessed and scored each video. For each resident, GEARS scores from the simulator tasks were compared with the GEARS scores in the operating room to assess the transferability of basic robotic skills from the simulator to the operating room.

Sample size calculation and statistical analysis

Sample size was calculated using the G* Power 3.1 for Mac, which was downloaded from the website: <http://www.gpower.hhu.de>. A sample size of 14 residents was required to detect a correlation with an effect size of 0.7 between the GEARS scores of the simulator and the operating room, with a two-tailed α error probability of 0.05, and power (1- β error probability) of 0.80.

Descriptive data are presented in terms of number, percentages, medians, and interquartile range. Spearman's rank correlation was used to assess the association between da Vinci Simulator scores on the simulator tasks with the GEARS scores from the operating room while performing bladder mobilization or urethro-vesical anastomosis. The Wilcoxon-Mann-Whitney test was performed to assess scoring differences between groups. The Wilcoxon signed-rank test was performed to assess scoring differences between the first and last attempt on the simulator tasks. Correlation between the simulator global scores and the GEARS scores on the videos of a given exercise was analyzed using the mixed linear model method to account for repeated measures [21]. All tests were two-sided with a significance threshold of 5%. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina).

Results

Five expert robotic surgeons were recruited to establish competency cut-offs for each of the 9 tasks on the da Vinci Surgical Skills Simulator. Their mean Global scores with standard deviations for the da Vinci Simulator tasks 1 through 9 were: 83.0 ± 22.8 ; 73.2 ± 19.4 ; 86.6 ± 11.9 ; 57.8 ± 28.1 ; 84.2 ± 25.2 ; 73.6 ± 18.4 ; 79.4 ± 24.6 ; 55.6 ± 24.3 ; 68.8 ± 27.8 , respectively. Therefore, the following competency cut-offs were established for the da Vinci Simulator Global scores: 61, 54, 75, 29, 59, 55, 55, 31, and 41 for tasks 1 through 9, respectively. The study included 14 urology residents. Seven residents were randomized to Group A and 7 residents were randomized to Group B. There were no significant differences between Groups A and B in terms of baseline characteristics (Table 1). Furthermore, there was no significant differences between both groups in terms of the mean number of practice sessions needed to reach competency (2 vs 1.8 practice sessions, for Groups A and B, respectively ($p=0.7$; Table 1). During this longitudinal study, 8 residents became senior residents (Post-Graduate Year 4-5) and performed the urethro-vesical anastomosis during Robotic Prostatectomy.

Comparison of performances between both groups across all simulation tasks

For both groups of residents, total Global scores obtained from the da Vinci Simulator correlated with the total GEARS scores for all 9 tasks ($p < 0.01$; Table 2). However, there was no significant difference between Groups A and B in their best Global scores from the da Vinci Simulator across all 9 tasks ($p=0.05$; Table 3). Similarly, when the GEARS tool was used to assess blinded videos of residents practicing on the da Vinci Simulator, there was no significant difference between the two groups ($p=0.05$; Table 3). When analyzing the Global scores of all residents pooled together, there were significant improvements from the first to the last attempt in the following tasks: Peg Board level 2 Match Board level 2, Needle Targeting, Ring and Rail level 2, Suture Sponge level 2, Energy Dissection level 1, and Ring Walk level 3 (all $p < 0.05$; Table 4). However, these did not translate to significant improvements in the total GEARS scores on the Simulator videos (all $p > 0.05$; Table 4).

Comparison of performance between the two groups in the operating room

There were no significant differences between Groups A and B in their total GEARS scores from their first three operating room procedures. The median total GEARS scores for the bladder mobilization were 19.9 (IQR: 19.7-20.8) and 20.0 (IQR: 19.9-21.2) for Groups A and B, respectively ($p=0.431$), and for the urethro-vesical anastomosis were 20.4 (IQR: 18.6-21.1) and 20.0 (IQR: 19.3-20.3) for Groups A and B, respectively ($p=0.774$) (Table 4). Similarly, there were no significant differences in the GEARS domains between the two groups (Table 4).

Comparison of performances between residents and attending urologists (experts) using GEARS in the operating room

GEARS was used to compare the performance of bladder mobilization and urethro-vesical anastomosis in the operating room. There was no significant difference in the total GEARS scores between attending urologists and residents during bladder mobilization (21.3 vs 19.9; $p>0.05$) and urethro-vesical anastomosis (20.8 vs 20.4; $p>0.05$).

Correlation between residents' simulator performance and performance in the operating room

For each simulator task, residents' Global scores were compared with their total GEARS scores in the operating room (Table 6). There was no significant correlation between the Global scores across 9 simulator tasks and total GEARS scores during bladder mobilization. However, a positive correlation was found between simulator task 4 (Ring and Rail 2) and urethro-vesical anastomosis ($\rho=0.74$, $p=0.037$) (Table 6).

For each simulator task, residents' total GEARS scores were correlated with their total GEARS scores from the operating room (Table 7). There was no significant correlation between the total GEARS scores on all 9 tasks and total GEARS scores during bladder mobilization. However, significant positive correlations were noted between total GEARS scores during simulator Task 4 (Ring and Rail 2; $\rho=0.86$, $p=0.007$) and Task 6 (Suture Sponge 2; $\rho=0.90$, $p=0.002$) with total GEARS scores during urethro-vesical anastomosis (Table 7).

Correlation between GEARS individual domains on the simulator and in the operating room.

GEARS individual domains were compared between the simulator and the operating room and multiple positive correlations were noted (Figure 1). Bimanual dexterity domain during the urethro-vesical anastomosis correlated significantly with bimanual dexterity domain during Tasks 5 (Dots and Needles 1) and 6 (Suture Sponge 2) ($\rho=0.74$; $p=0.037$ and $\rho=0.76$; $p=0.028$) (Figure 1).

The efficiency domain during Task 4 (Ring and Rail 2) correlated significantly with efficiency domain during urethro-vesical anastomosis ($\rho=0.90$, $p<0.001$) (Figure 1). In addition, the efficiency domain during Task 7 (Energy Dissection 1) correlated significantly with efficiency domain during bladder mobilization ($\rho=0.62$, $p=0.033$) (Figure 1). The force sensitivity domain during Task 4 (Ring and rail 2) and Task 5 (Dots and Needles 1) correlated with force sensitivity domain during bladder mobilization ($\rho=0.58$; $p=0.047$ and $\rho=0.65$; $p=0.022$). Finally, the robotic control domain during Task 6 (Suture Sponge 2) correlated with the same domain during urethro-vesical anastomosis ($\rho=0.81$, $p=0.015$).

Discussion

The first objective of the study was to establish competency cut-offs. Using the norm-referenced method, the following competency cut-offs for da Vinci Simulator Global scores were established: 61, 54, 75, 29, 59, 55, 55, 31, and 41 for tasks 1 through 9, respectively. These

Global scores are out of 100. The lowest competency cut-off score was 29 for task 4 (Ring and Rail 2), indicating that this was one of the most difficult tasks on the simulator. Interestingly, this was also the task that its Global and GEARS scores correlated significantly with the GEARS scores during the urethro-vesical anastomosis (Tables 6 and 7). At this point, urology residents were randomized to either practice 3 sessions (Group A) or practice till they reached competency (Group B). Our hypothesis was that residents would need at least 3 sessions to achieve competency. Surprisingly, both groups achieved competence after 2 and 1.8 sessions respectively (Table 1). Therefore, it was not surprising that both groups had no significant differences in their total Global scores from the da Vinci Simulator and total GEARS scores on the simulator and in the operating room (Table 3). Perhaps, if the study was designed with one arm as no practice or practice one session only, there could have been significant differences seen between the two groups. Other explanations for the lack of difference between the two groups could be that the competency cut-offs were set too low. When compared with a previous study including more robotic surgeons from different surgical specialties, it seems that the mean Global scores for the expert cohort in the current study is similar to mean Global scores from intermediates [19]. This could be explained by the fact that at the time of the study robotic surgery was still novel and expertise were still early, therefore, an expert performance at the time would equate to an intermediate performance in this era. Therefore, it is possible that if higher competency cut-offs were chosen, there may have been a significant difference between the two groups. Furthermore, the possibility of the sensitivity of the assessment tools used failed to detect smaller differences that would translate into a significant difference. Nevertheless, practice did in fact significantly improve Global scores for most tasks practiced (Table 4).

The second objective of the study was to compare the Global scores from the simulator to the blinded CSATS evaluations of the simulator tasks using the GEARS tool. This was an important step in assessing transferability of competency in basic robotic skills from the simulator to the operating room since there are no equivalency to the Global scores in the operating room. In fact, there was statistically significant correlation of the Global scores from the da Vinci simulator to the GEARS scores for each of the 9 tasks ($p < 0.01$; Table 2). Therefore, GEARS could be used to objectively assess performance on the da Vinci Simulator for the above 9 tasks.

The third objective of the study was to demonstrate transferability of basic robotic skills from the simulator to the operating room. In the present study, da Vinci simulator Global score and GEARS total scores from Ring and Rail 2 task significantly correlated with residents' GEARS total scores during urethro-vesical anastomosis. In addition, GEARS total scores from Ring and Rail 2 and Suture sponge level 2 tasks significantly correlated with residents' GEARS total scores during urethro-vesical anastomosis demonstrating the transfer of basic robotic skills from these two simulator tasks to the operating room (Table 2 and 3). When individual domains of the GEARS tool were analyzed, more positive correlations were found. Good bimanual

dexterity on the simulator tasks Dots and needle and Suture sponge level 2 tasks translated into good bimanual dexterity while performing the urethro-vesical anastomosis (Figure 1). Furthermore, the efficiency domain in Energy Dissection translated into good efficiency score when mobilizing the bladder, indicating that this task mimics several components of the bladder mobilization step in terms of use of energy and control of bleeding vessels using both robotic arms.

These findings have not been reported previously. In fact, the transferability of basic robotic skills was investigated previously with mostly negative findings. When 11 attending surgeons were assessed on the simulator and in the operating room, there was no correlation found between the Global scores of the simulator to the GEARS scores in the operating room [22]. However, in that study, there was no GEARS assessment of the simulator tasks, whereas in the present study, Global scores from the simulator were first correlated to the GEARS scores on the simulator then the GEARS scores from the simulator were correlated to the GEARS in the operating room. Another reason why there was no correlation could have been that multiple robotic surgeries were compared (thoracic, gynecologic, and urologic), whereas in the present study, only Robotic Prostatectomy was used. In another study, when medical students were randomized to practice to reach competency vs one practice session, there was no difference in the GEARS scores between both groups when they performed urethro-vesical anastomosis on a porcine model [23]. This is similar to the present study where there was no difference in the GEARS scores between the two groups of residents. Finally, one study demonstrated that a lengthy training protocol of training for one hour for 60 sessions on a simulator task (Tubes 3) resulted in faster performance of the urethro-vesical anastomosis [24]. However, in that study, there was no other objective assessment of technical skills other than measuring time taken to perform the anastomosis [24].

This study is not without limitations. The first limitation is the lack of control group without practice on the da Vinci Simulator. However, it was not ethical to deny robotic-naïve residents the opportunity to practice on the simulator prior to performing bladder mobilization during Robotic Prostatectomy. Therefore, it was decided to randomize residents to two groups one with practice for 3 sessions and the second group with practice till they reached competency. Furthermore, from a practical and patient safety stand-point it would have been difficult to accommodate a console naïve resident in a time restricted environment of the operating room. The strengths of this study are the fact that residents were randomized to two groups of training, in which all residents were console naïve at the time of enrollment capturing the very first attempts on the simulator, bladder mobilization and urethra-vesical anastomosis. The second limitation was the small sample size of residents and attending urologists despite recruiting all urology residents and all robotic surgeons in the urology training program. The third limitation was that this study was conducted during the initial robotic experience of the urology department and the mean Global scores of the experts in this cohort was similar to intermediates in previous

publications. This could explain the lower competency cut-offs and the lack of statistically significant difference between the GEARS scores of the experts and the trainees in the operating room. Another explanation for the lack of statistical difference is that the GEARS tool as assessed by lay CSATS was not sensitive enough to detect small differences in performance in the operating room. The fourth limitation was that residents were not re-assessed on the da Vinci simulator as they progressed in their residency training. Therefore, it is possible that senior residents' robotic skills improved since their assessment on the simulator during the simulator phase of the study. This could explain the lack of correlation of total GEARS scores on tasks other than Ring and Rail 2 with total GEARS scores during urethro-vesical anastomosis. Nevertheless, this study correlated simulator Global scores with total GEARS scores for all 9 tasks. In addition, the da Vinci simulator Global score and total GEARS scores from Ring and Rail 2 task significantly correlated with residents' total GEARS scores during urethro-vesical anastomosis, demonstrating the transfer of basic robotic skills from the simulator to the operating room.

Conclusions

Objective assessments of urology residents on the da Vinci Surgical Skills Simulator tasks Ring and Rail 2 and Suture Sponge correlated with their objective assessments of bladder mobilization and urethro-vesical anastomosis. Therefore, basic robotic skills could be transferred from the simulator to the operating room.

References

1. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin.* 2014; 64:252–71.
2. Heidenreich A, Bastian PJ, Bellmunt J, Bolla M, Joniau S, Van der Kwast T, et al. EAU guidelines on prostate cancer. Part 1: screening diagnosis and treatment of clinically localized disease. *Eur Urol.* 2011; 59:61–71.
3. Abboudi H, Khan MS, Guru KA, Froghi S, de Win G, Van Poppel H, et al. Learning curves for urological procedures: a systematic review. *BJU Int.* 2014;114: 617–29.
4. Gumus E, Boylu U, Turan T, Onol FF. The learning curve of robot assisted radical prostatectomy. *J Endourol.* 2011; 25:1633–637.
5. O'Malley PJ, Van Appledorn S, Bouchier-Hayes DM, Crowe H, Costello AJ. Robotic radical prostatectomy in Australia: initial experience. *World J Urol.* 2006; 24:165–170.
6. Sooriakumaran P, John M, Leung R, et al. A multi-institutional study of 3794 patients undergoing robotic-assisted radical prostatectomy shows the learning curve is not as short as previously thought. *Eur Urol Suppl.* 2011; 10:1569–9056.
7. Newcomb L, Bradley M, Troung T, Tang M, Combstock B, Li YJ, et al. Correlation of Virtual Reality Simulation and Dry Lab Robotic Technical Skills. *J Minim Invasive Gynecol.* 2018;25(4):689-696
8. Whiteburst S, Lockrow E, Lendvay T, Propst A, Dunlow SG, Rosemeyer CJ, et al. Comparison of Two Simulation Systems to Support Robotic-Assisted Surgical Training: A Pilot Study (Swine Model). *J of Minim Invasive Gynecol.* 2015;23;5: 483-488.
9. Volpe A, Ahmed K, Dasgupta P, Foccaro V, Novara G, Van der Poel, et al. Pilot Validation Study of the European Association of Urology Robotic Training Curriculum. *Eur Urol.* 2015; 68:292–299.
10. Stegemann AP, Ahmed K, Syed JR, Rehman S, Ghani K, Antorino R, et al. Fundamental skills of robotic surgery: a multi-institutional randomized controlled trial for validation of a simulation-based curriculum. *Urology.* 2013; 81:767–774.
11. Foell K, Finelli A, Yasufuku K, Waddel TK, Pace KT, Honey RJ, et al. Robotic surgery basic skills training: Evaluation of a pilot multidisciplinary simulation- based curriculum. *Can Urol Assoc J* 2013; 7:430–434.
12. Dioun S, Fleming N, Munsell M, Lee J, Ramirez PT, Soliman PT. Setting Benchmarks for the New User: Training on the Robotic Simulator. *JSLs:* 2017;21;4: 1-5
13. Goh A, Goldfarb D, Sander J, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol* 2012;187(1):247–252.
14. Ghani KR, Miller DC, Linsell S, Brachulis A, Lane B, Sarle R, et al. Measuring to Improve: Peer and Crowd-sourced Assessments of Technical Skill with Robot-assisted Radical Prostatectomy. *Eur Urol.* 2016 Apr;69(4):547-550.
15. Goldenberg MG, Goldenberg L, Grantcharov TP: Surgeon Performance Predicts Early Continence After Robot-Assisted Radical Prostatectomy. *J Endourol,* 2017. 31(9): 858-863

16. Powers MK, Boonjindasup A, Pinsky M, Dorsy P, Maddox M, Su LK, et al. Crowdsourcing assessment of surgeon dissection of renal artery and vein during robotic partial nephrectomy: a novel approach for quantitative assessment of surgical performance. *J Endourol* 2016; 30(4):447–452.
17. Ghani K, Miller D, Linsell S, Brachullis A, Lane B, Sarle R, et al. Measuring to Improve: Peer and Crowd-sourced Assessments of Technical Skill with Robot-assisted Radical Prostatectomy. *Eur Urol*; 2016;69;547-550
18. Holst D, Kowalewski TM, White LW, Brand TC, Harper JD, Sorenson MD, et al. Crowd-sourced assessment of technical skills: an adjunct to urology resident surgical simulation training. *J Endourol* 2014;29(5):604–609
19. Alzahrani T, Haddad R, Alkhayal A, Delisle J, Drudi L, Gotlieb W, et al. Validation of the da Vinci Surgical Skill Simulator across three surgical disciplines: A pilot study. *Can Urol Assoc J*. 2013;7(7-8): E520-9.
20. George S, Haque MS, Oyebode F. Standard setting: Comparison of two methods. *BMC Med Educ*. 2006; 6:46.
21. Hamlett A, Ryan L, Serrano-Trespacios P, Wolfinger R. Mixed models for assessing correlation in the presence of replication. *Journal of the Air & Waste Management Association*. 2003;53(4):442-50.
22. Mills JT, Houggen HY, Bitner D, Krupski TL, Schenkman NS. Does Robotic Surgical Simulator Performance Correlate with Surgical Skill? *J Surg Educ*. 2017;74(6):1052-1056.
23. Vargas M, Moawad G, Denny K, Happ L, Misa NY, Marguilies S, et al. Transferability of Virtual Reality, Simulation-Based, Robotic Suturing Skills to a Live Porcine Model in Novice Surgeons: A Single-Blind Randomized Controlled Trial. *J of Minim Invasive Gynecol*, 2017;24(3):420-425.
24. Shim JS, Noh TI, Kim JY, Pyun JH, Cho S, Oh MM, et al. Predictive Validation of a Robotic Virtual Reality Simulator: The Tube 3 module for Practicing Vesicourethral Anastomosis in Robot-Assisted Radical Prostatectomy. *Urology*. 2018; 122:32-36

Figures and Tables

Fig. 1. Spearman rho correlations for individual GEARS domains between the simulator and the operating room for each simulator task. BM: bladder mobilization; UVA: urethro-vesical anastomosis.

	GEARS domain									
	Depth Perception		Bimanual Dexterity		Efficiency		Force Sensitivity		Robotic Control	
	BM	UVA	BM	UVA	BM	UVA	BM	UVA	BM	UVA
Task 1	-0.33 p=0.299	0.19 p=0.651	-0.27 p=0.391	0.05 p=0.910	0.10 p=0.745	0.07 p=0.867	0.15 p=0.633	-0.10 p=0.823	-0.04 p=0.897	0.55 p=0.160
Task 2	-0.06 p=0.846	-0.05 p=0.9108	0.03 p=0.931	0.36 p=0.385	0.12 p=0.713	0.50 p=0.207	0.04 p=0.897	0.55 p=0.160	-0.01 p=0.966	0.29 p=0.493
Task 3	-0.38 p=0.217	0.07 p=0.8665	-0.39 p=0.208	0.33 p=0.419	0.38 p=0.217	0.62 p=0.102	-0.14 p=0.6646	-0.05 p=0.910	-0.44 p=0.152	0.50 p=0.207
Task 4	-0.48 p=0.118	-0.21 p=0.610	0.57 p=0.051	0.67 p=0.071	0.52 p=0.085	0.90 p=0.002	0.58 p=0.047	-0.05 p=0.911	-0.15 p=0.633	0.21 p=0.610
Task 5	0.03 p=0.914	0.62 p=0.102	-0.14 p=0.665	0.74 p=0.037	0.39 p=0.208	0.64 p=0.086	0.65 p=0.022	0.38 p=0.352	0.00 p=1.000	-0.24 p=0.570
Task 6	-0.45 p=0.138	-0.33 p=0.419	0.15 p=0.648	0.76 p=0.028	0.54 p=0.071	0.62 p=0.102	-0.09 p=0.779	0.31 p=0.456	0.05 p=0.880	0.81 p=0.015
Task 7	0.07 p=0.829	0.43 p=0.2894	0.12 p=0.713	0.57 p=0.139	0.62 p=0.033	0.17 p=0.693	0.36 p=0.255	0.24 p=0.570	-0.34 p=0.286	0.10 p=0.823
Task 8	-0.45 p=0.138	-0.17 p=0.693	-0.01 p=0.966	0.38 p=0.352	0.07 p=0.829	0.24 p=0.570	0.24 p=0.457	-0.45 p=0.260	-0.20 p=0.527	0.40 p=0.320
Task 9	-0.06 p=0.846	0.38 p=0.352	-0.36 p=0.226	0.19 p=0.651	-0.04 p=0.897	0.48 p=0.233	0.06 p=0.845	0.14 p=0.736	-0.08 p=0.795	0.29 p=0.493

-1	-0.6	-0.2	0.0	0.2	0.6	1
----	------	------	-----	-----	-----	---

Legend for Spearman rho coefficients

Variables	Group A (n=7)	Group B (n=7)	p
Age (mean)	28	28	0.9
Gender			
Male	5 (71%)	5 (71%)	1.0
Female	2 (29%)	2 (29%)	
Pre-study number of laparoscopic cases as first assistant (mean)	16.7	13.9	0.8
Practice sessions needed to reach competency in all 9 tasks (mean)	2	1.86	0.7

Task	Task name	Correlation coefficient	p
1	Pegboard level 2	0.75	<0.001
2	Match board level 2	0.60	<0.001
3	Needle targeting	0.64	<0.001
4	Ring & rail level 2	0.49	0.002
5	Dots & needles level 1	0.50	<0.001
6	Suture sponge level 2	0.61	<0.001
7	Energy & dissection level 1	0.63	<0.001
8	Ring walk level 3	0.34	0.009
9	Tubes	0.58	<0.001

Table 3. Simulator global and GEARS total scores for each of the tasks for groups A and B			
	Simulator global scores		
Task #	Group A	Group B	p
1	96 (89–98)	96 (94–97)	0.934
2	79 (67–83)	82 (78–86)	0.559
3	94 (85–100)	97 (87–99)	0.593
4	82 (74–93)	74 (52–83)	0.175
5	82 (63–87)	85 (74–93)	0.209
6	81 (64–86)	74 (71–92)	0.730
7	85 (85–91)	95 (80–96)	0.419
8	67 (47–81)	73 (50–90)	0.473
9	63 (56–71)	74 (53–92)	0.400
	Simulator GEARS total scores		
Task #	Group A	Group B	p
1	20.0 (19.7–21.6)	21.5 (20.4–21.9)	0.383
2	20.0 (19.0–20.3)	20.0 (19.6–21.2)	0.902
3	19.8 (18.3–20.9)	19.9 (19.3–20.2)	1.000
4	19.3 (19.0–19.7)	19.2 (19.2–19.5)	0.805
5	18.2 (16.9–20.0)	18.7 (17.9–19.1)	0.620
6	19.3 (18.7–20.0)	19.2 (18.2–19.4)	0.318
7	20.5 (19.7–20.7)	19.6 (19.1–20.6)	0.259
8	18.9 (18.4–19.5)	19.5 (18.6–19.8)	0.620
9	18.5 (18.2–19.6)	18.9 (18.0–19.6)	0.900

Data are presented as medians (interquartile range).

Table 4. Differences in global scores and GEARS score from the first attempt to the last attempt for each simulator task for all residents in both groups					
	Task	Global score median (IQR)	p	GEARS median (IQR)	p
1	Pegboard level 2	4.5 (3, 11)	0.001	0.95 (-0.09, 1.63)	0.127
2	Match board level 2	9.5 (1, 19)	0.028	0.78 (-0.51, 2.65)	0.147
3	Needle targeting	10 (1, 17)	0.005	0.66 (-0.55, 1.76)	0.217
4	Ring & rail level 2	8 (-3, 25)	0.049	0.32 (-0.35, 0.80)	0.417
5	Dots & needles level 1	3.5 (-4, 17)	0.095	0.26 (-0.75, 1.95)	0.326
6	Suture sponge level 2	11 (-1, 22)	0.047	0.40 (-0.44, 2.45)	0.127
7	Energy & dissection level 1	3 (-2, 9)	0.042	-0.18 (1.33, 0.82)	1
8	Ring walk level 3	15.5 (0, 27)	0.023	0.41 (-0.95, 1.89)	0.426
9	Tubes	6.5 (1, 12)	0.235	0.31 (-0.36, 0.75)	0.241

IQR: interquartile range.

Operation room procedure	GEARS domain	Group A	Group B	p
Bladder mobilization	Depth perception	4.1 (4.0–4.2)	4.0 (3.9–4.1)	0.876
	Bimanual dexterity	4.1 (3.9–4.2)	4.2 (4.1–4.3)	0.343
	Efficiency	3.7 (3.5–3.9)	3.6 (3.3–3.8)	0.639
	Force sensitivity	4.0 (3.9–4.1)	4.2 (4.1–4.4)	0.149
	Robotic control	4.3 (4.0–4.4)	4.2 (4.2–4.6)	0.432
	Total	19.9 (19.7–20.8)	20.0 (19.9–21.2)	0.431
Urethro-vesical anastomosis	Depth perception	4.0 (3.8–4.2)	4.1 (3.9–4.1)	1.000
	Bimanual dexterity	4.1 (3.7–4.3)	4.0 (3.8–4.1)	1.000
	Efficiency	3.9 (3.2–3.9)	3.5 (3.4–3.6)	0.571
	Force sensitivity	4.0 (3.9–4.2)	4.1 (3.9–4.1)	1.000
	Robotic control	4.4 (4.0–4.5)	4.5 (4.1–4.5)	0.786
	Total	20.4 (18.6–21.1)	20.0 (19.3–20.3)	0.774

Data are presented as medians (interquartile range).

	Task	Bladder mobilization		Urethro-vesical anastomosis	
		rho	p	rho	p
1	Pegboard level 2	-0.31	0.322	0.18	0.668
2	Match board level 2	-0.10	0.753	0.61	0.108
3	Needle targeting	-0.47	0.122	0.18	0.668
4	Ring & rail level 2	0.08	0.812	0.74	0.037
5	Dots & needles level 1	-0.07	0.829	0.24	0.570
6	Suture sponge level 2	-0.34	0.276	0.60	0.120
7	Energy & dissection level 1	-0.24	0.445	0.35	0.396
8	Ring walk level 3	-0.40	0.199	0.10	0.822
9	Tubes	-0.14	0.672	0.32	0.435

Table 7. Correlation between simulator total GEARS scores and operating room total GEARS scores

Task	Task	Bladder mobilization		Urethro-vesical anastomosis	
		rho	p	rho	p
1	Pegboard level 2	-0.27	0.391	0.17	0.693
2	Match board level 2	-0.11	0.729	0.29	0.493
3	Needle targeting	-0.41	0.183	0.07	0.867
4	Ring & rail level 2	0.15	0.649	0.86	0.007
5	Dots & needles level 1	0.25	0.430	0.50	0.207
6	Suture sponge level 2	-0.22	0.485	0.90	0.002
7	Energy & dissection level 1	-0.08	0.795	0.29	0.493
8	Ring walk level 3	-0.24	0.457	0.00	1.000
9	Tubes	-0.37	0.236	0.31	0.456