A high-fidelity, virtual reality, transurethral resection of bladder tumor simulator: Validation as a tool for training

Jonathan Moore[1]; Stewart Whalen[2]; Neal Rowe[3]; Jason Lee[4]; Michael Ordon[4]; Andrea Lantz-Powers[2]

[1]Department of Urology, Mayo Clinic, Phoenix, AZ, United States; [2]Department of Urology, Dalhousie University, Halifax, NS, Canada; [3]Division of Urology, Department of Surgery, University of Ottawa and the Ottawa Hospital Research Institute, Ottawa, ON, Canada; [4]Division of Urology, Department of Surgery, University of Toronto, Toronto, ON, Canada

Corresponding author: Dr. Stewart Whalen, Department of Urology, Dalhousie University, Halifax, NS, Canada; stewart.whalen@dal.ca

***

Abstract

Introduction: Simulation-based training is used to help trainees learn surgical procedures in a safe environment. The objective of our study was to test the face, content, and construct validity of the transurethral resection of bladder tumor (TURBT) module built on the Simbionix TURP Mentor simulator.

Methods: Participants performed five standardized cases on the simulator. Domains of the simulator were evaluated on a five-point Likert scale to establish face and content validity. Construct validity was assessed through the simulator's built-in scoring metrics, as well as video recordings of the simulator screen and an anonymized view of participants' hands and feet, which were evaluated using an objective structured assessment of technical skills (OSATS) tool.

Results: Ten experienced operators and 15 novices participated. Face validity was somewhat acceptable (mean realism 3.8/5±1.03 standard deviation [SD]; mean appearance 4.1/5±0.57), as was content validity, represented by simulation of key steps (mean 3.9±0.57). The simulator failed to achieve construct validity. There was no difference in mean simulator scores or OSATS

scoring between experienced operators and novices. Novices significantly improved their mean simulator scores (305.9 vs. 332.4, p=0.006) and OSATS scoring (15.8 vs. 18.1, p=0.001), while 87% felt their confidence to perform TURBT improved. Overall, 92% of participants agreed that the simulator should be incorporated into residency training.

Conclusions: Our study suggests a role for the TURBT module of the Simbionix TURP Mentor simulator as an introduction to TURBT for urology trainees. Strong support was found from both experienced operators and novices for its formal inclusion in resident education.

Introduction

An essential milestone in urology residency training is proficiency in transurethral resection of bladder tumors (TURBT). Learning TURBTs is associated with a significant learning curve.[1] Operator experience has been shown to affect outcomes including complication rates, some of which are associated with high morbidity or mortality.[1-3] For example, gross hematuria progressing to transfusion has been shown to be more common with trainees.[3] Recurrence rates of bladder tumors as well as presence of detrusor muscle in resected samples (a key criteria for staging of tumors) are both dependent on level of experience.[1,2] Even senior residents face significant challenges learning TURBT as they tend to take on more difficult cases and are more likely to be involved in complications than junior residents.[3]

An emerging alternative to traditional surgical training is high fidelity simulation-based training. A range of methods including bench models, cadavers and virtual reality simulators have been used.[4] These new technologies have the potential to offer training opportunities in a safe environment. Simulation training is being studied across a number of specialties including urology. Trainers have been developed for percutaneous nephrolithotomy, ureteroscopy, robotic pelvic surgery, and transurethral resection of prostate (TURP).[5-10]

As a relatively short cystoscopic procedure, TURBT is well suited to simulation-based training. It can also be built on the same platform and with similar software as a TURP simulator. To date, there has been no successfully validated virtual reality training simulators for TURBT. Two previous bench models, the Bristol TURBT Trainer and the Simbla TURBT simulator, have demonstrated face, content and construct validity.[11,12] These simulators both used physical models complete with irrigation, cautery, and camera systems.

This study aims to evaluate the TURBT module on the Simbionix TURP Mentor simulator (Simbionix LTD, Airport City, Israel), for face, content and construct validity.

Methods

Residents and attending urologists were recruited from an academic Canadian urology training program. Participants performed a standardized simulation protocol on the TURBT simulator. An initial questionnaire was administered to all participants prior to beginning the simulations.

This was used to obtain demographics, level of training and experience with cystoscopy, TURP, TURBT, and simulation-based training. Attitudes towards simulation were investigated along with a self-assessment of cystoscopic skills.

To start, participants had a 15-minute introduction, orientation and calibration session which included a warm-up period where they were able to use the simulator on a low difficulty case. Participants then performed five TURBT simulations and were evaluated. Each attempt required the participant to perform a complete cystoscopy and resect three separate papillary or nodular tumors. The simulator did not require fulguration of the resection bed or edges for cancer control. Fulguration was required for hemostasis. Four cases were available on the simulator with increasing levels of difficulty. A standardized progression through the two "moderately" difficult cases was done by all participants (case 2, 3, 2, 3, and 3) to provide some variety and progression of difficulty. After each attempt, participants were given feedback on performance by the simulator.

A post-training questionnaire was administered to all participants and included qualitative information on participant attitudes and their assessment of the simulator as well as their opinions on the role of the TURBT simulator in urology training. Both the pre- and post-training questionnaires were developed by all members of the research team in an iterative process with collaboration from the department of urology research coordination office.

The simulator was assessed for face, content and construct validity in concordance with definitions published in the literature.[13] Face and content validity were evaluated using "experienced" operator assessment of the simulator only. Based on criteria established in previous studies, experienced operator participants were defined as having performed greater than 50 real life TURBTs.[14,15] Using a five-point Likert scale, participants evaluated multiple domains of the simulator including depth of resection, resectoscope movement and fluid management. Higher numbers indicated more favorable assessment. Thresholds were set a-priori as a mean score of <3.0, 3.0-4.0, and >4.0 for "unacceptable", "somewhat acceptable", and "acceptable", respectively. Schout et al. conducted a review of criteria for validation within their study of a TURP simulator and found a lack of consistency within the literature.[14] They did find however, that the majority of studies used similar thresholds and adapted them to a 5, 7 or 10-point Likert scale.

Construct validity was assessed using two methods of participant evaluation. These consisted of simulator generated scores and independent blinded urologist review of video recordings of the simulator screen and an anonymized view of participants' hands and feet. Simulator generated scores of participant performance were available for five domains: resection completion, bleeding control, safety, economy, and visualization. The maximum possible score was 370. Economy was measured by the cumulative path length of resection (mm) and time to complete resection (seconds). The cumulative path length was calculated as the distance the loop traveled while the control pedal for cut/coagulation was engaged. The simulator halted after it

detected a perforation and provided an overall score of zero on that attempt. However, individual scores for all domains were still calculated by the simulator and available for analysis. The simulator falsely awarded better scores for resection time and cumulative path length on attempts with perforation due to early termination of the trial. As such, for our analysis we assigned scores of 0 for resection time and cumulative path length during attempts in which a perforation occurred. Other domain scores were calculated up until the point of perforation by the simulator and summed by the investigators manually to yield the overall score for that attempt.

Videos of the simulator screen and an anonymized view of participants' hands and feet were evaluated by two independent urologists using an objective structured assessment of technical skills (OSATS) tool (Figure 1). The tool was adapted from a validated assessment tool for surgical residents.[16] Domains assessed were respect for tissue, time and motion, instrument handling, flow of operation, knowledge of procedure, overall performance, and safety. The maximum possible score was 25. Videos of the first and last attempts for each participant were scored. Evaluators were blinded to experienced/novice status of the participant, but not to the attempt number.

Data was analyzed using SPSS (IBM SPSS Statistics, Version 24). Categorical variables were compared using Fisher's exact test. For evaluation of continuous variables, Mann Whitney U test was used for group comparisons and Wilcoxon Sign-Rank Test for paired data. An alpha of 0.05 was set for significance of all statistical tests. Study design was reviewed and approved by the local Research Ethics Board.

Results

Twenty-five participants completed the study – ten experienced operators and fifteen novices (Table 1). Novices were a heterogenous group of residents from post-graduate year (PGY) 1-5 with varied cystoscopic and TUR experience. One resident was classified as experienced based on the definition of having performed greater than 50 real life TURBTs. The other nine experienced operators were staff urologists encompassing a variety of subspecialties including endourology, oncology, transplant, and reconstruction. Six participants had prior experience on the Simbionix TURP Mentor simulator while only one had used the TURBT module previously.

Face validity was evaluated using two questions looking at the realism and appearance of the simulation (Figure 2). Mean scores for experienced operators were "somewhat acceptable" for realism (3.8, SD 1.03) and "acceptable" for appearance (4.1, SD 0.57).

Overall assessment and individual procedure components were evaluated to test for content validity (Figure 2). Overall assessment of content validity was "somewhat acceptable" for experienced operators (3.9, SD 0.57). Experienced operators rated camera movement (4.4, SD 0.7) and movement in 3D space (4.4, SD 0.52) as "acceptable." Simulation of bleeding control (3.5, SD 1.18), scope feedback (3.2, SD 1.03), and resectoscope movement (3.9, SD 0.74) were "somewhat acceptable." "Unacceptable" domains were fluid management (2.8, SD 1.09), control of bladder distention (2.8, SD 0.63) and resection depth (2.7, SD 1.16).

Construct validity combined both simulator generated scores and OSATS scoring. There was no difference in overall mean simulator scores aggregated across all five attempts between experienced and novice operators (330.9 vs. 333.7, p=0.892, Table 2). This was also the case when attempts with bladder perforation were analyzed separately (256.1 vs 269.2, p=0.739). Rate of perforation was not significantly different between experienced operators and novices (22% vs. 28%, p=0.643). The five individual categories assessed showed no significant differences between experienced operators and novices.

A comparison of the mean novice simulator scores between the second and fifth attempt (both case number 3) showed a significant improvement (p=0.006) from 305.9 (SD 50.6, range 210-357) to 332.4 (SD 28.6, range 256-363). This was not observed in the experienced group where scores were 314.9 (SD 30.7, range 261-349) and 325.4 (SD 33.8, range 248-359) for the second and fifth attempt, respectively (p=0.386).

OSATS scoring of videos showed no significant difference between total scores for experienced operators and novices on their first (17.6 vs 15.8, p=0.255) and fifth attempts (19.8 vs 18.1, p=0.345, Table 3). There was a non-significant trend towards better scores in the experienced group. Experienced operators were not significantly more likely to receive a "pass" by the evaluators than novices on the first (70% vs 46.7%, p=0.270) or fifth attempt (60% vs 53.3%, p=0.7). There were no failures in the experienced operator group. Novices recorded failure rates of 20% and 6.7% during their first and fifth attempts respectively. Novice operators were significantly more likely to receive an unsafe grade on their first and fifth attempts (46.7% vs. 0%, p=0.0026 for both). Evaluation of total OSATS scores from the first to fifth attempt showed significant improvements for both experienced operators (17.6 vs 19.8, p=0.004) and novices (15.8 vs 18.1, p=0.001). The magnitude of the change was similar for experienced (+2.2 points) and novice (+2.3 points) operators.

Cancelling and restarting an attempt due to technical reasons was an uncommon event that occurred in 9/134 (6.7%) of trials. These were due to malfunction of the scope calibration during the attempt making resection impossible.

Written narrative qualitative assessment by participants showed several common themes. Participants found the simulator to have good "fidelity," "movement," and a "realistic resectoscope." Most participants, however, had concerns about how the simulator assessed depth of resection. Many reported the "feel" of the resection and the fluid management were weaknesses. Participants noted that the simulator lacked the ability to remove tumor chips from the bladder and a realistic simulation for fulguration of resection margins.

During the post-simulation questionnaire, novices were asked if they felt their confidence to perform a real life TURBT had improved. Data showed that 20% "strongly agreed", 67% "agreed" and the remaining 22% were neutral about the statement. Forty-four percent of all

participants "strongly agreed" and 48% "agreed" that the TURBT simulator should be incorporated into the urology residency training curriculum.

Discussion

Simulation is at the forefront of modern surgical education. Our study showed novice learners were able to improve confidence in their TURBT skills after performing only 5 cases on the simulator (87% "agree" or "strongly agree"). Furthermore, there was a significant improvement in total simulator scores for novices that was not seen in the experienced group, indicating progression in novices' ability to complete the simulated tasks. This was also reflected in novice OSATS scores which improved in the same interval. Some of the improvement in scoring may be due to acclimatization to the simulator, as improvement was observed in OSATS scores for experienced operators as well.

Based on the predefined definitions of face and content validity, our study showed the TURBT module built on the Simbionix TURP Mentor simulator met criteria for both at a somewhat acceptable level. However, it failed to meet criteria for construct validity. Using both the built-in evaluation metrics and the OSATS evaluation tool, the simulator was unable to differentiate between experienced and novice operators. One possible explanation is that the simulator is not difficult enough to create a separation. Specifically, one of the most challenging aspects of a TURBT is resecting at the correct depth to obtain muscle in the specimen without unduly thinning the bladder or causing a perforation. Participants consistently reported poor simulation of depth of resection.

The categorization of participants as experienced or novice, based on the definition of greater or fewer than 50 real life TURBTs, was arbitrary and self-reported by participants. Overall participant number was relatively low and this study may not have had statistical power to detect differences between groups. Failure and lack of safety was only seen in the novice group. In this regard, OSATS evaluation was able to identify those participants who were the most unsafe with statistical significance.

While there have been many validation studies of TURP simulators, there is minimal data in the literature on TURBT simulator validation. The UroTrainer (Karl Storz GmbH, Tuttlingen, Germany), a high fidelity endourological simulator with both TURP and TURBT modules, has been evaluated in several studies. Reich and colleagues initially published on the development of the trainer without formal validation.[17] A subsequent study failed to achieve face and content validity when evaluated in the setting of an international urology conference.[14] The study authors, however, made important points concerning the lack of established tools and cutoffs for validity testing. A second group looked at the UroTrainer, but randomized participants to groups with and without photodynamic diagnostics.[18] They found improvement in performance for novice trainees over 5 training rounds in both groups. The novice and experienced groups were exposed to different training modules and no method of basic validation was used (face, content, etc.). More recently, the UroTrainer was studied in a group of medical students with no

cystoscopic experience.[19] A significant improvement in performance was seen after a period of training. No experienced group was used; therefore, face, content and construct validity could not be assessed. Another simulator, the Simbla TURBT, is bench model-based and uses synthetic bladders as well as real resectoscopes and irrigation, differentiating it from the virtual reality simulator used in this study. De Vries et al. showed this simulator to have acceptable realism (face validity), simulation of steps (content validity) and construct validity.[12] Construct validity was established in their study by demonstrating overall performance of novices rated lower compared with intermediates and experts.

Areas for improvement for this and future simulators have been revealed in this study. These include an improved scoring metric to better differentiate operators and level of skill and higher fidelity simulation of bladder distention and depth of resection. An ideal simulator should change the thickness of the bladder based on distention, have visual representation of resection into the bladder mucosa/muscularis, and provide feedback on the presence of muscle in the specimen. Despite the noted deficiencies, the clear majority (92%) "agreed" or "strongly agreed" that the simulator should be part of the urology residency curriculum.

One of the main limitations of our study is the small sample size (n=25). Expanding the study to a larger number of participants would be reasonable based on our results. Additionally, changes in OSATS evaluations from first to fifth attempt may have been subject to bias as there was blinding to participant experience and identity, but not to attempt number. Evaluators may have preferentially attributed higher scores to later attempts. The clinical correlation of an improvement in a single raw score to better real-life skills was beyond the scope of this study. It is not yet possible to assess a trainee's ability using the simulator scores alone. However, the improved subjective confidence and OSATS scores suggest benefit from the simulator and supports its use as a tool for training.

Conclusions

The TURBT simulator built on the Simbionix TURP Mentor showed face and content validity at a somewhat acceptable level but failed to demonstrate construct validity. Novices demonstrated improved scores on the simulator as well as improved confidence after only 5 simulated cases. Both experienced and novice participants agreed that the simulator should be incorporated into the urology residency curriculum. Based on our study, we suggest this simulator could be used as an introduction to TURBT for novices to gain experience and skills in a safe environment. Further progression in skills could be accelerated by the addition of real time feedback from staff urologists or senior residents. There is currently not enough evidence to recommend this simulator for the evaluation of participants' operative skill. Further research is required to correlate simulator performance to clinical TURBT performance in real life.

References

1. Mariappan P, Zachou A, Grigor KM, et al. Detrusor muscle in the first, apparently complete transurethral resection of bladder tumour specimen is a surrogate marker of resection quality, predicts risk of early recurrence, and is dependent on operator experience. *Eur Urol.* 2010;57:843-49.
2. Jancke G, Rosell J, Jahnson S. Impact of surgical experience on recurrence and progression after transurethral resection of bladder tumour in non-muscle-invasive bladder cancer. *Scand J Urol.* 2014;48:276-83.
3. Nieder AM, Meinbach DS, Kim SS, et al. Transurethral bladder tumor resection: intraoperative and postoperative complications in a residency setting. *J Urol.* 2005;174:2307-09.
4. Aydin A, Shafi AM, Shamim Khan M, et al. Current Status of Simulation and Training Models in Urological Surgery: A Systematic Review. *J Urol.* 2016;196:312-20.
5. Veneziano D, Smith A, Reihsen T, et al. The SimPORTAL fluoro-less C-arm trainer: an innovative device for percutaneous kidney access. *J Endourol.* 2015;29:240-45.
6. White MA, Dehaan AP, Stephens DD, et al. Validation of a high fidelity adult ureteroscopy and renoscopy simulator. *J Urol.* 2010;183:673-77.
7. Kumar A, Smith R, Patel VR. Current status of robotic simulators in acquisition of robotic surgical skills. *Curr Opin Urol.* 2015;25:168-74.
8. Kumar PV, Gomes MP, Davies BL, et al. A computer assisted surgical trainer for transurethral resection of the prostate. *J Urol.* 2002;168:2111-14.
9. Gomes MP, Barrett AR, Timoney AG, et al. A computer-assisted training/monitoring system for TURP structure and design. *IEEE Trans Inf Technol Biomed.* 1999;3:242-51.
10. Brewin J, Ahmed K, Khan MS, et al. Face, content, and construct validation of the Bristol TURP trainer. *J Surg Educ.* 2014;71:500-05.
11. Shamim Khan M, Ahmed K, Gavazzi A, et al. Development and implementation of centralized simulation training: evaluation of feasibility, acceptability and construct validity. *BJU Int.* 2013;111:518-23.
12. de Vries AH, van Genugten HG, Hendrikx AJ, et al. The Simbla TURBT Simulator in Urological Residency Training: From Needs Analysis to Validation. *J Endourol.* 2016;30:580-87.
13. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc.* 2003;17:1525-29.
14. Schout BM, Bemelmans BL, Martens EJ, et al. How useful and realistic is the uro trainer for training transurethral prostate and bladder tumor resection procedures? *J Urol.* 2009;181:1297-1303

15. Mishra S, Kurien A, Ganpule A, et al. Face and content validity of transurethral resection of prostate on Uro Trainer: is the simulation training useful? *J Endourol.* 2010;24:1839-43.

16. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997;84:273-78.

17. Reich O, Noll M, Gratzke C, et al. High-level virtual reality simulator for endourologic procedures of lower urinary tract. *Urology.* 2006;67:1144-48.

18. Kruck S, Bedke J, Hennenlotter J, et al. Virtual bladder tumor transurethral resection: an objective evaluation tool to overcome learning curves with and without photodynamic diagnostics. *Urol Int.* 2011;87:138-42.

19. Neumann E, Mayer J, Russo GI, et al. Transurethral Resection of Bladder Tumors: Next-generation Virtual Reality Training for Surgeons. *Eur Urol Focus.* 2019;5:906–11.

Figures and Tables

Fig. 1. Objective structured assessment of technical skills (OSATS) tool.

Generic skill							
Respect for tissue	**1** Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments	2	**3** Careful handling of tissue but occasionally caused inadvertent damage	4	**5** Consistently handled tissues appropriately with minimal damage		
Time & motion	**1** Many unnecessary moves	2	**3** Efficient time/motion but some unnecessary moves	4	**5** Economy of movement and maximum efficiency		
Instrument handling	**1** Repeatedly makes tentative or awkward moves with instruments	2	**3** Competent use of instruments although occasionally appeared stiff or awkward	4	**5** Fluid moves with instruments and no awkwardness		
Flow of operation & forward planning	**1** Frequently stopped operating or needed to discuss next move	2	**3** Demonstrated ability for forward planning with steady progression of operative procedure	4	**5** Obviously planned course of operation with effortless flow from one move to the next		
Knowledge of specific procedure	**1** Deficient knowledge, Needed specific instruction at most operative steps	2	**3** Knew all important aspect of the operation	4	**5** Demonstrated familiarity with all aspects of the operation		
Rate the participant's performance (circle)	Fail		Borderline Pass		Pass		
Participant is able to safely perform this operation independently (circle)			Y		N		

Fig. 2. Overall assessment and individual procedure components were evaluated to test for content validity.
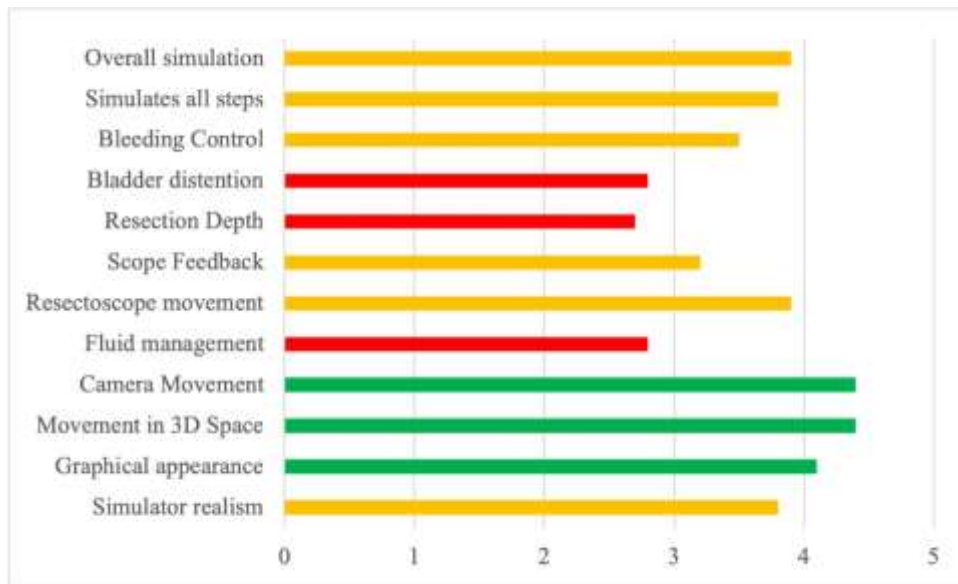
Table 1. Participant baseline demographics		
	Experienced	**Novice**
Total (n)	10	15
Female (%)	1 (10%)	6 (40%)
Left-handed (%)	1 (10%)	2 (13.3%)
Sim experience		
TURP	2 (20%)	4 (26.7%)
TURBT	0 (0%)	1 (6.7%)
TURBT experience		
0 cases	0 (0%)	8 (53%)
<5 cases	0 (0%)	3 (20%)
5−15 cases	0 (0%)	1 (6.7%)
15−25 cases	0 (0%)	2 (13.3%)
25−50 cases	0 (0%)	1 (6.7%)
>50 cases	10 (100%)	0 (0%)

TURP: transurethral resection of the prostate; TURBT: transurethral resection of the bladder tumor.

Table 2. Results of experienced and novice participants based on built-in metrics on the simulator			
Measure	**Experienced**	**Novice**	**p**
Metric (SD)	n=10	n=15	
Resection	101.1 (16.3)	102.8 (14.7)	0.849
Bleeding control	68.5 (1.8)	68.7 (1.4)	0.935
Safety	67.4 (5.6)	68.7 (1.4)	0.935
Economy	67.1 (2.6)	65.9 (2.5)	0.196
Visualization	26.8 (3.19)	27.75 (2.9)	0.643
Total scores			
Attempt 1	352 (11.4)	340.1 (34.2)	0.531
Attempt 2	314.9 (30.7)	305.9 (50.6)	0.807
Attempt 3	349.1 (15.1)	354.9 (16.8)	0.397
Attempt 4	300.1 (51.7)	310.1 (44.7)	0.892
Attempt 5	325.4 (33.8)	332.4 (28.6)	0.605
Overall	330.9 (24.6)	333.7 (19.9)	0.892
Perforation rate (%)	22%	28%	0.643*
Cumulative path length (mm)	923.4 (290.0)	1059.4 (236.5)	0.216
Resection time (sec)	247.2 (71.4)	285.6 (62.5)	0.311

*Fisher's exact test. P-values were calculated using Mann Whitney U test, unless otherwise indicated. SD: standard deviation.

Table 3. OSATS scores on first and fifth attempt							
		Attempt 1			Attempt 5		
		Experienced	**Novice**	**p**	**Experienced**	**Novice**	**p**
Pass/Fail							
	Pass	7 (70%)	7 (46.7%)	0.270	6 (60%)	8 (53.3%)	0.700
	Borderline	3 (30%)	5 (33.3%)		4 (40%)	6 (40%)	
	Fail	0	3 (20%)		0	1 (6.7%)	
Safety							
	Yes	7 (70%)	7 (46.7%)	0.026	7 (70%)	7 (46.7%)	0.026
	Borderline	3 (30%)	1 (6.7%)		3 (30%)	1 (6.7%)	
	No	0	7 (46.7%)		0	7 (46.7%)	
Total Score (SD)		17.6 (2.5)	15.8 (4.8)	0.255	19.8 (2.7)	18.1 (4.7)	0.345

p=0.004*

p=0.001*

*Wilcoxon rank sum. P-values were calculated using Fisher's exact test, unless otherwise indicated.