Objective Assessment and Standard Setting for Basic Flexible Ureterorenoscopy Skills Among Urology Trainees Using Simulation-Based Methods

Mitchell Goldenberg, MBBS,¹ Michael Ordon, MD, FRCSC,¹ John R. D'A Honey, MD, FRCSC,¹
Sero Andonian,² and Jason Y. Lee, MD, MHPE, FRCSC³

Abstract

Objective: To objectively assess the performance of graduating urology residents performing flexible ureterorenoscopy (fURS) using a simulation-based model and to set an entrustability standard or benchmark for use across the educational spectrum.

Methods: Chief urology residents and attending endourologists performed a standardized fURS task (ureterorenoscopy and repositioning of stones) using a Boston Scientific[®] Lithovue ureteroscope on a Cook Medical[®] URS model. All performances were video-recorded and blindly scored by both endourology experts and crowdworkers (C-SATS) using the Ureteroscopic Global Rating Scale, plus an overall entrustability score. Validity evidence supporting the scores was collected and categorized. The Borderline Group (BG) method was used to set absolute performance standards for the expert and crowdsourced ratings.

Results: A total of 44 participants (40 chief residents, 4 faculties) completed testing. Eighty-three percent of participants had performed >50 fURS cases at the time of the study. Only 47.7% (mean score 12.6/20) and 61.4% (mean score 12.4/20) of participants were deemed “entrustable” by experts and crowdworkers, respectively. The BG method produced entrustability benchmarks of 11.8/20 for experts and 11.4/20 for crowdworker ratings, resulting in pass rates of 56.9% and 61.4%.

Conclusion: Using absolute standard setting methods, benchmark scores were set to identify trainees who could safely carry out fURS in the simulated setting. Only 60% of residents in our cohort were rated as entrustable. These findings support the use of benchmarks to earlier identify trainees requiring remediation.

Keywords: assessment, education, minimally invasive surgery, urology, surgery simulation

Introduction

THE MOVE TO COMPETENCY-BASED MEDICAL EDUCATION (CBME) requires a shift in not only curricular design, but the frequency and significance of assessments as well.¹ This shift focuses on outcome-based assessments (e.g., Milestones in United States, Competency By Design in Canada),² and a systematic and structured approach to assessments in CBME is critical for educators to make defensible high-stakes decisions about trainee competency.³

The valid assessment of technical skills is paramount to surgical training. As evidence has pointed to their impact not only on educational outcomes in residency, but also patient outcomes,⁴ there has been a recent surge in the creation of

objective assessment tools and simulation-based educational intervention designs and implementation.⁵ The use of such assessments in the context of a competency-based curriculum can be structured using the Entrustable Professional Activities (EPAs) framework that relies on iterative collection of assessment data from a trainee over the course of their training to ensure that a trainee is able to safely carry out a given clinical task or surgical procedure, at the time of graduation.^{6,7} Although these frameworks have been established in many jurisdictions,⁸ we still lack benchmarks or standards in these assessments that identify those trainees who have reached this level of “entrustability.”⁶ Of equal importance, benchmarks are needed in this context to accurately identify trainees who require further training and remediation.⁹

¹Division of Urology, Department of Surgery, St. Michael's Hospital, University of Toronto, Toronto, Canada.

²Division of Urology, McGill University Health Centre, McGill University, Quebec, Canada.

³Division of Urology, Department of Surgery, University Health Network—Toronto General Hospital, University of Toronto, Toronto, Canada.

The use of standard setting methodology in procedural assessments has been explored in urology previously.^{10,11} Benchmarks in the context of skill assessments can be set using relative or absolute methods, with the former using the performance of an index cohort (i.e., experts' scores when completing the task) and the latter using a predefined set of criterion that reflects the purpose of the assessment.¹² Although both are acceptable when for use in low-stakes or formative assessments, it is generally accepted that absolute methods are more appropriate when making high-stakes or summative decisions.¹³

In urologic surgery, flexible ureterorenoscopy (fURS)¹⁴ is a core requisite skill and it is expected that graduating residents will be competent in the skills necessary to perform fURS.¹⁵ One of the tools used to assess fURS skill is a procedural-specific global rating scale (GRS), adapted from the Objective Structured Assessment of Technical Skills (OSATS) rubric,¹⁶ published by Matsumoto and colleagues.¹⁷ Although recent literature has used this rating tool to set norm-referenced standards (based on attending surgeon performance),¹⁸ efforts have yet to be undertaken to set benchmarks of entrustability, to determine when trainees are safe to independently perform fURS.

In this study, we aimed to objectively assess the basic fURS skills of graduating residents using a procedure-specific ureteroscopic GRS, to set an entrustability standard suitable for use in both formative and summative assessments. We hypothesize that by taking a rigorous approach to standard setting using an accepted absolute method, we will set a defensible benchmark for entrustability assessments in fURS.

Methods

After receiving institutional research ethics board approval (REB#16-034), chief residents (final year of residency) from urology programs across Canada were recruited to participate in the study. In addition, four academic endourologists, who are experts in fURS (all were in practice for >5 years), were recruited. All participants were assessed performing a standardized simulation-based fURS task (described below).

Standardized task

All participants completed a standardized task using the Cook Medical[®] URS model, an acrylic and polycarbonate model that includes a dual collecting system.¹⁴ The task consisted of participants advancing a flexible ureteroscope (Boston Scientific Lithovue[®]) through a ureteral access sheath into the simulated kidney. They were asked to perform a complete diagnostic ureterorenoscopy, which involved fully inspecting all eight calices in the model, then relocating two previously placed lower caliceal stones to the ipsilateral upper calix using a 1.3F nitinol tipless basket. The participants were aided by a surgical assistant controlling the basket, but this assistant only acted when instructed to by the participant and was not permitted to give any advice or guidance. We did not aim to assess the skill of laser lithotripsy. The same assistant was used by all participants. No specific warm-up was allowed before testing, but all participants were oriented to the flexible ureteroscope, the basket, and the training model.

Data collection

Three raters, with both content expertise (endourologists) and education backgrounds, independently and blindly scored the performance using a modified version of the ureteroscopic GRS,¹⁷ with possible scores ranging from 4 to 20 points (Fig. 1). In addition, raters were asked to provide an overall score of 1–3, based on an appraisal of task “entrustability” (Fig. 1). These raters had all previously used the rating instrument in the study setting and were oriented with the purpose of this study before completing their assessments.

In addition to ratings from expert surgeons, the videos collected were sent to C-SATS, Inc.[®],¹⁹ a web-based platform that crowdsources procedural videos to a network of “crowd-workers,” providing rapid high-volume assessments of technical skills. These crowd-workers were recruited anonymously through the C-SATS, Inc.[®] platform, and over multiple iterations have been shown in other urologic procedures to provide accurate and reliable ratings of technical skill.²⁰ This additional set of ratings were conducted to understand how ratings of fURS technical skill set by crowd-workers (nonurologists) compares with those of expert surgeons, and to provide additional validity evidence for the entrustability standard set. Crowd-workers were provided with training material before participating, as well as didactic information regarding the procedure itself, including clinical indications for fURS. Crowd-workers were asked to use the ureteroscopic GRS and overall entrustability score to blindly rate the videos. C-SATS provided overall entrustability ratings for each case as the proportion of crowd-workers who selected each category (1–3). Raters were oriented to the assessment method and rating scale and were paid a fee for each video scored.

Data analysis

Descriptive statistics were calculated for demographic and performance scores, with overall pass–borderline–fail decisions coded as ordinal data. Two separate GRS scores were calculated for each participant using (1) the three expert ratings and (2) the crowd-worker rating. The sum of means for the individual GRS domain scores provides the final GRS scores for both sets of raters. The overall entrustability score allowed the cohort to be stratified into three groups, for both expert and crowdsourced scores. Majority rule was used to determine final participant status in the case of rater disagreement regarding entrustability. Inter-rater reliability was calculated for both expert and crowd-worker GRS scores using the intraclass coefficient, and for overall entrustability score using Cronbach's alpha. Internal consistency of the ureteroscopic GRS was calculated by calculating agreement between domains of the instrument. To set a competency benchmark, we used the Borderline Group (BG) method of standard setting. We carried out two standard setting procedures for the expert and crowdsourced ratings, and additionally applied the two calculated standards to each set of GRS ratings. Receiver operating characteristic (ROC) curves were calculated to identify variables that accurately predict a trainee's pass/fail status, including time to complete task and previous fURS experience.²¹ Two-tailed *p*-values <0.05 were considered statistically significant. All statistical analyses were performed using SPSS v24 (NY).

Please circle the number corresponding to the candidate's performance in each category, irrespective of level of training.				
Respect for Tissue	1 Scope frequently pushed into urothelial wall. Used unnecessary force with basket.	2	3 Scope occasionally pushed into urothelial wall. Careful handling of basket for the most part.	4 No trauma to urothelial wall with scope. Consistent and careful handling of basket.
Time and Motion	1 Many unnecessary moves.	2	3 Made some unnecessary moves but time more efficient.	4 No unnecessary moves and time <u>is</u> maximized.
Instrument Handling	1 Needed to repeatedly attempt basketing of stone.	2	3 Able to basket stone within first few tries. Occasional awkward movement.	4 Able to basket with fluid motion and no awkwardness.
Adaptability and strategic movements	1 Displays no ability to adapt to difficulty; demonstrated little to no trouble shooting abilities	2	3 Was able to demonstrate some adaptability to difficult situations; eventually tried different strategies or approaches to overcome difficulty	4 Adapted easily to difficulty; demonstrates timely ability to try new strategic approaches to overcome difficulty

FIG. 1. Ureteroscopy global rating scale.

Overall Rating:

How would you rate the overall performance of this surgeon?

PASS	BORDERLINE PERFORMANCE	FAIL
can perform safely and independently	could perform with some supervision	cannot perform safely nor independently

Results

For a 2-year study period, 40 graduating residents representing 11 urology programs across Canada participated in the study. In addition, four attending endourologists participated in the study as experts (Table 1). Notably, all participants felt they would be competent in fURS at the completion of residency. All 44 performances were blindly rated by the experts. C-SATS provided 1069 ratings by 270 individual crowd-workers, for a 7.1 hour period. Crowd-workers completed a mean of 38.3 ratings per video.

According to the expert ratings, 21 (47.7%, mean score 15.3) participants were deemed "entrustable," 14 (31.8%, mean score 11.8) were "borderline," and 9 (20.5%, mean score 7.3) were found not safe to complete the task independently (Table 2). Crowd-workers rated 27 (61.4%, mean score 13.4) participants as "entrustable," 8 (18.2%, mean score 11.4) as "borderline," and 7 (15.9%, mean score 10.4) as not safe to independently complete the task. Expert and crowd-worker entrustability ratings moderately correlated (0.407, $p=0.01$). There were inconsistencies regarding en-

trustment status of individual participants, with only three participants rated as "fail," 5 as "borderline," and 16 as "pass" by both groups.

Mean time to task completion among participants was 323 seconds. Strong inverse correlations were seen between the time taken to complete the task and expert rating (Pearson's correlation = -0.915) and crowdsourced ratings (Pearson's correlation = -0.811). Reported fURS case volume did not correlate with either expert (Spearman's rho = -0.070) or crowd ratings (Spearman's rho = -0.146), but level of training (trainee vs faculty) correlated with both task time (339 seconds vs 158 seconds, $p=0.05$) and expert ratings (mean 12.6 vs 16.8, $p=0.04$).

Expert raters demonstrated excellent intraclass correlation (0.912, $p<0.01$). Regarding overall entrustability decisions among expert raters, agreement was excellent (Cronbach's alpha = 0.973). Inter-rater reliability of crowd ratings was not provided by C-SATS, but they provided a correlation calculation between two of their own expert raters and crowd-workers (Spearman's rho = 0.85). However, inter-rater agreement regarding overall entrustability decisions for

TABLE 1. PARTICIPANT DEMOGRAPHICS

	N (%)
Program	
University of British Columbia	3 (6.8)
University of Alberta	4 (9.1)
University of Manitoba	4 (9.1)
University of Western Ontario	4 (9.1)
McMaster University	7 (15.9)
University of Toronto	6 (13.6)
University of Ottawa	3 (6.8)
Queen's University	3 (6.8)
McGill University	5 (11.4)
University of Montreal	3 (6.8)
Dalhousie University	1 (2.3)
Flexible ureterorenoscopy experience	
None/minimal	0 (0)
<10 cases	0 (0)
10–25	1 (2.3)
25–50	8 (18.2)
50–100	16 (36.4)
>100 cases	19 (43.2)
Simulation-based training in flexible ureterorenoscopy	
Yes	13 (29.5)
No	27 (61.4)
Unknown	4 (9.1)
Competent at flexible ureterorenoscopy?	
Yes	44 (100)
No	0 (0)

crowd-workers was much more varied, with the percentage of raters agreeing on whether a trainee was entrustable ranging from 18% to 96%, borderline from 4% to 49%, and unsafe from 0% to 45% across participants. The internal consistency of the ureteroscopic GRS as rated by expert surgeons was found to be excellent (Cronbach's $\alpha = 0.911$), similar to crowd ratings (Cronbach's $\alpha = 0.979$).

The BG method was applied to both expert and crowd-worker ratings. Using expert ratings, the mean score of borderline participants was 11.8/20 (Fig. 2a). Using this benchmark, 19 (43.1%) participants failed and 25 (56.9%)

participants passed the assessment. Using crowdsourced ratings, the mean of borderline participants was 11.4/20 (Fig. 2b). Applying this cutoff score to the crowdsourced assessment scores, 17 (38.6%) participants failed and 27 (61.4%) passed the assessment. When comparing the agreement in pass/fail status between these two cohorts of scores, discrepancy in decisions was seen in eight cases (18.2%), with five participants passing the C-SATS assessment who failed the experts' assessment, and three failing the C-SATS assessment who passed the experts' scoring.

ROC curves were calculated to test the accuracy of both time-to-task completion and previous fURS experience as predictors of participants reaching the standard. Time-to-task completion was highly predictive of whether a participant would achieve both the expert (area under the curve [AUC] = 0.977, SE 0.02, $p = 0.00$) and the crowdsourced (AUC = 0.913, SE = 0.04, $p = 0.00$) standards. However, previous fURS experience was not a significant predictor of either standard, with both AUC values <0.500.

Discussion

The study data provide evidence toward the setting of entrustability benchmarks in technical skill for fURS, using a method of setting standards appropriate for both low- and high-stakes assessments. Our findings used two sets of assessors, expert surgeons and crowdsourcing, to evaluate the technical ability of graduating urology trainees for a 2-year period. Interestingly, these data indicate that despite all participants' perception of their own competency in this task at the time of graduation, only 56.9% to 61.4% would have passed this simulation-based assessment.

This study adds to a growing body of work evaluating the use of simulation-based tasks for high-stakes assessment.² Despite investigation into the use of simulation for the training and evaluation of technical skills in endoscopic prostate surgery,²² laparoscopy,^{11,23} robotics,²⁴ and endourology, there remains a lack of implementation of these assessments for summative decision-making in residency.² Emerging evidence supports the use of simulation-based assessment to predict clinical performance,^{5,25} and this kind of literature supports the validity and defensibility of using

TABLE 2. CORRELATIONS BETWEEN PERFORMANCE RATINGS

	Expert rating (three raters)	C-SATS, Inc.® rating (270 crowd-workers)	Correlation statistic	
	Mean (SD)	Mean (SD)		
Ureteroscopic GRS				
Respect for tissue	3.2 (0.8)	3.2 (0.4)	0.67	$p < 0.00$
Time and motion	3.0 (1.1)	3.0 (0.5)	0.77	$p < 0.00$
Instrument handling	3.1 (1.1)	3.0 (0.6)	0.82	$p < 0.00$
Adaptability	3.2 (1.0)	3.3 (0.5)	0.78	$p < 0.00$
Total	12.5 (3.7)	12.5 (1.9)	0.83	$p < 0.00$
	n (%)	n (%)		
Overall entrustability score				
Can perform safely and independently	21 (47.7)	27 (61.4)	0.407	$p = 0.01$
Can perform with some supervision	14 (31.8)	8 (18.2)		
Cannot perform safely	9 (20.5)	7 (15.9)		

GRS = global rating scale.

FIG. 2. Standard setting using the Borderline Group method (a) expert ratings (b) crowd-sourced ratings (C-SATS). dark gray line, fail; light gray line, borderline; black line, pass.

scores from these assessments for credentialing purposes in urologic surgery.

These data also add to the literature investigating the application of standard setting to procedural technical skill assessment.¹² The standards set in this study are in congruence with Norcini's principles, in that a sound methodology was applied, raters were appropriately selected and oriented, and the standards are credible and realistic.²⁶ Using the BG Method to set benchmarks for the ureteroscopic GRS is appropriate given the inclusion of a borderline score on the overall entrustability scale, and since it is an "absolute" standard setting method, it is appropriate to use for high-stakes assessments.^{27,28}

This study used crowdsourcing as an additional source of validity evidence for both performance scores, and the credibility of the standards set. These data support the use of crowdsourcing as a means of providing rapid high-volume assessments of technical skills, and the standard setting method applied to both sets of ratings produced similar benchmarks (11.4/20 vs 11.8/20). However, it should be noted that although the standards created using expert and GRS were similar, the distribution of scores was not, with crowd-workers much less willing to score performances as either very good or very poor. As a result, the reported excellent reliability of crowdsourcing assessments in this context, and in previous study,^{29,30} may reflect crowd-worker's

tendency to gravitate toward average scores (3/5) on GRS scoring rubrics; suggesting perhaps such crowdsourcing assessments are not ideal for high-stakes complex procedural assessments.

This study has limitations that should be noted when interpreting its findings. First, the relatively small sample size, as well as the single performance captured per participant, may limit the generalizability of our findings. However, the wide distribution in GRS scores and entrustability ratings may indicate that we adequately captured a sufficient range of technical skill levels, reflective of the variability in ability of graduating residents, to support our benchmark's validity. Second, the BG method, although well documented in the literature as a means of creating absolute standards, does not allow cut-point adjustments to mitigate errors in sensitivity (minimizing false-negative or false-positive rate).¹³ This ability to modify the standard based on the purpose of the assessment is associated with the Contrasting Groups method, which was not possible given the assessment tool used in our study. Third, the relatively new LithoVue® flexible ureteroscope may have introduced an unintentional bias and skewed the performance scores, and thereby the entrustability ratings, in the negative direction. It is also possible that the low pass rate among these graduating residents is simply a result of being "rusty," as FURS cases are often delegated to junior- or mid-level residents. In addition,

these performance standards were set based on objective performance scores for a simulated fURS task. Further study to ensure similar findings correlate to performance in the clinical setting are warranted. Another limitation to our study is that we did not analyze the “communication” abilities of each participant in performing the task; the ability to properly and effectively instruct the basketing assistant could have been influential on fURS performance. Finally, one must acknowledge the inherent bias of using crowd-sourced assessments to evaluate surgical skill. The overall similarities in both ratings and the standard set provide additional evidence for the utility of the C-SATS platform for assessing basic surgical skills.

Conclusion

Using absolute standard setting methods, benchmark scores were set to identify trainees who could safely carry out fURS in the simulated setting. Only 60% of residents in our cohort were rated as entrustable. These findings support the use of benchmarks to earlier identify trainees requiring remediation.

Acknowledgments

Boston Scientific and Cook Medical gratuitously provided the LithoVue disposable flexible ureteroscopes and URS training model, respectively, for the study.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

No funding was received for this article.

References

- Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR; for the International CBME Collaborators. The role of assessment in competency-based medical education. *Med Teach* 2010;32:676–682.
- Szasz P, Grantcharov TP, Sweet RM, et al. 8th Annual American College of Surgeons Accredited Educational Institutes (ACS-AEI) Consortium Simulation-based summative assessments in surgery. *Surgery* 2016;160:528–535.
- MacNeily A. The trouble with technical competence: Defining it, measuring it and achieving it! *Can Urol Assoc J* 2014;8:1–2.
- Fecso AB, Szasz P, Kerezov G, Grantcharov TP. The effect of technical performance on patient outcomes in surgery: A systematic review. *Ann Surg* 2016. DOI: 10.1097/SLA.0000000000001959
- Brydges R, Hatala R, Zendejas B, Erwin PJ, Cook DA. Linking simulation-based educational assessments and patient-related outcomes. *Acad Med* 2015;90:246–256.
- Pugh D, Cavalcanti RB, Halman S, et al. Using the entrustable professional activities framework in the assessment of procedural skills. *J Grad Med Educ* 2017;9:209–214.
- Holmboe ES, Call S, Ficalora RD. Milestones and competency-based medical education in internal medicine. *JAMA Intern Med* 2016;176:1601–1602.
- Swing S. The Urology Milestone Project 2015:1–42.
- Rethans JJ, Norcini JJ, Barón-Maldonado M, et al. The relationship between competence and performance: Implications for assessing practice performance. *Med Educ* 2002;36:901–909.
- Raison N, Ahmed K, Fossati N, et al. Competency based training in robotic surgery: Benchmark scores for virtual reality robotic simulation. *BJU Int* 2017;119:804–811.
- Lee JY, Andonian S, Pace KT, Grober E. Basic Laparoscopic Skills Assessment Study: Validation and standard setting among Canadian urology trainees. *J Urol* 2017;197:1539–1544.
- Goldenberg MG, Garbens A, Szasz P, Hauer T, Grantcharov TP. Systematic review to establish absolute standards for technical performance in surgery. *Br J Surg* 2017;104:13–21.
- Downing SM, Tekian A, Yudkowsky R. RESEARCH METHODOLOGY: Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teach Learn Med* 2006;18:50–57.
- Blankstein U, Lantz AG, D'A Honey RJ, Pace KT, Ordon M, Lee JY. Simulation-based flexible ureteroscopy training using a novel ureteroscopy part-task trainer. *Can Urol Assoc J* 2015;9:331–335.
- Morrison KB, MacNeily AE. Core competencies in surgery: Evaluating the goals of urology residency training in Canada. *Can J Surg* 2006;49:259–266.
- Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;84:273–278.
- Matsumoto ED, Hamstra SJ, Radomski SB. A novel approach to endourological training: Training at the Surgical Skills Center. *J Urol* 2001;166:1261–1266.
- Aloosh M, Couture F, Fahmy N, Elhilali MM, Andonian S. Assessment of urology postgraduate trainees' competencies in flexible ureteroscopic stone extraction. *Can Urol Assoc J* 2018;12:52–58.
- Holst D, Kowalewski TM, White LW, et al. Crowd-Sourced Assessment of Technical Skills (C-SATS): Differentiating animate surgical skill through the wisdom of crowds. *J Endourol* 2015;29:1183–1188.
- Dai JC, Lendvay TS, Sorensen MD. Crowdsourcing in surgical skills acquisition: A developing technology in surgical education. *J Grad Med Educ* 2017;9:697–705.
- Szasz P, Bonrath EM, Louridas M, et al. Setting performance standards for technical and nontechnical competence in general surgery. *Ann Surg* 2017;266:1–7.
- Aydin A, Raison N, Khan MS, Dasgupta P, Ahmed K. Simulation-based training and assessment in urological surgery. *Nat Rev Urol* 2016;13:503–519.
- Sweet RM, Beach R, Sainfort F, et al. Introduction and validation of the American Urological Association Basic Laparoscopic Urologic Surgery skills curriculum. *J Endourol* 2012;26:190–196.
- Smith R, Patel V, Satava R. Fundamentals of robotic surgery: A course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. *Int J Med Robot* 2014;10:379–384.

25. Aghazadeh MA, Mercado MA, Pan MM, Miles BJ, Goh AC. Performance of robotic simulated skills tasks is positively associated with clinical robotic surgical performance. *BJU Int* 2016;118:475–481.
26. Norcini JJ, Shea JA. The credibility and comparability of standards. *Appl Meas Educ* 1997;10:39–59.
27. Kane MT, Crooks TJ, Cohen AS. Designing and evaluating standard-setting procedures for licensure and certification tests. *Adv Health Sci Educ Theory Pract* 1999;4:195–207.
28. Norcini JJ. Setting standards on educational tests. *Med Educ* 2003;37:464–469.
29. White LW, Kowalewski TM, Dockter RL, Comstock B, Hannaford B, Lendvay TS. Crowd-Sourced Assessment of Technical Skill: A valid method for discriminating basic robotic surgery skills. *J Endourol* 2015;29:1295–1301.
30. Powers MK, Boonjindasup A, Pinsky M, et al. Crowd-sourcing 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;30:447–452.

Address correspondence to:
Jason Y. Lee, MD, MHPE, FRCSC
Division of Urology
Department of Surgery
University Health Network—Toronto General Hospital
585 University Avenue PMB 11, Room 186
Toronto
M5G 2N2 Canada

E-mail: jason.lee2@uhn.ca

Abbreviations Used

AUC = area under the curve
 BG = Borderline Group
 CBME = competency-based medical education
 EPAs = Entrustable Professional Activities
 fURS = flexible ureterorenoscopy
 GRS = global rating scale
 OSATS = Objective Structured Assessment of Technical Skills
 ROC = receiver operating characteristic