Understanding and Assessing Nontechnical Skills in Robotic Urological Surgery: A Systematic Review and Synthesis of the Validity Evidence

Jethro CC Kwong, BSc,* Jason Y Lee, MD,† and Mitchell G Goldenberg, MBBS†

*Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; and †Division of Urology, Department of Surgery, University of Toronto, Toronto, Ontario, Canada

OBJECTIVE: Robotic urological surgery (RUS) has seen widespread adoption across institutions in the last decade. To match this rapid growth, it is imperative to develop a structured RUS curriculum that addresses both technical and nontechnical competencies. Emerging evidence has shown that nontechnical skills form a critical component of RUS training. The purpose of this review is to examine the validity evidence of available nontechnical skills assessment tools in RUS.

METHODS: A literature search of MEDLINE, EMBASE, and PsycINFO was conducted to identify primary articles using nontechnical skills assessment tools in RUS. Messick's validity framework and the Medical Education Research Study Quality Instrument were utilized to evaluate the quality of the validity evidence of the abstracted articles.

RESULTS: Of the 566 articles identified, 12 used nontechnical skills assessment tools in RUS. The metrics used ranged from self-assessment using global rating scales, to objective measures such as electroencephalography. The setting of these evaluations ranged from immersive and virtual reality-based simulators to live surgery.

CONCLUSIONS: Limited effort has been made to develop nontechnical skills assessment tools in RUS. Recently, there has been a shift from subjective to objective measures of nontechnical performance, as well as the development of assessments specific to RUS. However, the validity evidence supporting these nontechnical assessments is limited at this time, including their relationship to technical skills, and their impact on surgical outcomes. (J Surg Ed 76:193−200. © 2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: Urology residency training, Robotic surgery, Nontechnical skill assessment, Simulation, Education

COMPETENCIES: Interpersonal and Communication Skills, Practice-Based Learning and Improvement

INTRODUCTION

Robotic urological surgery (RUS) has been on the rise in the past decade, and is being widely adopted as the preferred approach to various operations including radical prostatectomy, partial nephrectomy, and pyeloplasty, among others.[1] As the global demand for RUS continues to increase, there is a growing need for a framework of RUS training and accreditation.[2] In the current competency-based system, trainees must progress through both technical and nontechnical milestones to develop proficiency in RUS. However, the design and implementation of a structured curriculum incorporating both elements is still in its infancy. Great attention has been paid to the development and validation of technical skill assessment tools, evident by the surge of novel simulators and metrics.[3] However, similar efforts in nontechnical domains have lagged behind.

The role of nontechnical skills has previously been described in multiple surgical disciplines and has been linked to patient outcomes and overall operative performance.[4–6] Several metrics have been developed to assess nontechnical performance, including Nontechnical Skills for Surgeons (NOTSS), Non-Technical Skills

Correspondence: Inquiries to Mitchell G Goldenberg, MBBS Division of Urology, Department of Surgery, University of Toronto, St. Michael's Hospital, 30 Bond Street, 16CC-056, Toronto, Ontario M5B 1W8, Canada; e-mail: mitchell.goldenberg@mail.utoronto.ca

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

 1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2018.05.009

(NOTECHS), and Observational Teamwork Assessment for Surgery (OTAS).[5,7,8] These assessment tools encompass three domains: cognitive skills (decision-making, planning, and situation awareness), personal resource factors (stress and fatigue), and social skills (communication, teamwork, and leadership).[9] However, the robotic surgical ecosystem presents unique challenges to surgeons, namely the dichotomous roles of the console surgeon and bedside assistant, and the lack of tactile feedback.[10] These complexities magnify the extent and breadth of the surgeon's cognitive workload and as such, necessitate a strong command of nontechnical skills specific to RUS.

The purpose of this systematic review was to survey the landscape of nontechnical skill assessment tools in RUS and to critically appraise the supporting validity evidence for these metrics. To adequately evaluate the validity of these assessment methods, a contemporary framework was used to structure the evidence abstracted from the included articles. As a secondary objective, this review assessed the quality of evidence presented in the included studies.

METHODS

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol.[11]

Eligibility Criteria

Articles that assessed nontechnical skills in RUS in trainees (medical students, residents, and fellows) and surgeons were eligible for inclusion. Nontechnical skills were defined as qualities outside of technical domains and include cognitive skills (decision-making, planning, and situation awareness), personal resource factors (stress and fatigue), and social skills (communication, teamwork, and leadership).[9] Articles were eligible regardless of publication status and study type. Editorials, opinion letters, case reports, reviews, and letters to the editor were excluded.

Information Sources

A search was conducted in Ovid MEDLINE, Embase Classic, PsycINFO, and the Cochrane Library on July 18, 2017.

Search

The search strategy has been previously described in detail,[3] which was designed to capture articles assessing both technical and nontechnical skill in RUS. Medical subject headings terms used in the search specific to nontechnical skills included "communication," "nontechnical skill," and "cognitive skill".

Study Selection

Titles and abstracts generated from the search were screened independently by two authors for full-text review. Any disagreements were resolved by consensus. Articles referenced in the included studies were eligible for inclusion.

Data Collection

The following data was extracted from the included articles: sample size, participant type, assessment used, setting of assessment, domains tested, rater type, and information relevant to the evaluation of study quality and validity evidence. Two authors reviewed the extracted data and any disagreements were thoroughly discussed.

Data Analysis (Quality of Literature and Validity Evidence)

The Medical Education Research Study Quality Instrument (MERSQI) was used to evaluate the quality of evidence in the included articles.[12] This tool appraises methodological quality of medical education research across the following domains: study design, sampling, type of data, validity of evaluation instrument, data analysis, and outcomes.

Messick's validity framework[13] was used to evaluate the validity evidence for judgments made using the nontechnical skill assessment tools in the included articles. The validity evidence was categorized into the following domains: content, response process, internal structure, relationships to other variables, and consequences of testing.

RESULTS

The search terms yielded 566 articles (Fig). Each article was screened according to the eligibility criteria and all eligible articles were crosschecked for article references. A total of 12 articles were selected for full text review, and are listed in Appendix A. Using Messick's validity framework, the evidence supporting nontechnical skill assessment tools in RUS are summarized in the Table.

NASA Task Load Index

The NASA Task Load Index (NASA-TLX) was originally developed by Hart and Staveland[14] for the aviation industry in 1988, and has since expanded into healthcare and surgery. This subjective assessment tool consists of six domains designed to capture an individual's mental

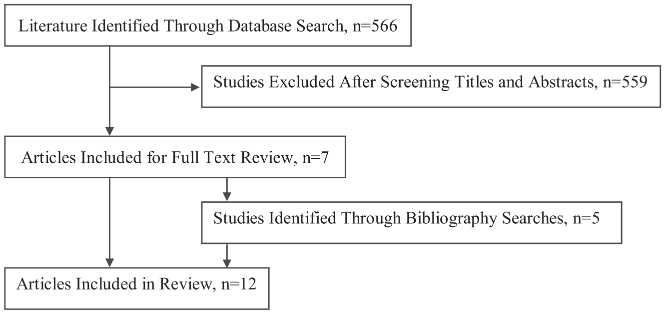

FIGURE. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart.

response to a given task, with a higher score representing higher cognitive workload. These domains include mental demand, physical demand, temporal demand, performance, effort, and frustration.

The NASA-TLX presents the strongest validity evidence for the assessment of nontechnical skills in RUS, and has demonstrated a broad range of applications. Klein et al. found that cognitive workload was reduced when performing identical tasks under stereoscopic vision compared to two-dimensional vision.[15] During the initial validation studies of the Mimic dV-Trainer (MdVT), Sethi et al. showed that experts had lower scores than novices in all domains except for temporal demand.[16] The NASA-TLX also demonstrated similar workload profiles when performing tasks on the MdVT and live porcine models,[17] supporting its use in simulation-based exercises for RUS training. Notably, Sessa et al. found comparable workload profiles between novices and experts when using the Xperience Team Trainer, a bedside assistant training platform to the MdVT.[18] In addition, the NASA-TLX has highlighted the importance of procedure-based training and mentorship in RUS curricula, which contributes to marked improvements in levels of cognitive distress, understanding, and engagement.[19,20] Lastly, it is the only assessment tool with supporting consequences evidence, demonstrating a strong association between cognitive workload and inattention blindness in the peripheral visual fields.[21]

Interpersonal and Cognitive Assessment for Robotic Surgery

The Interpersonal and Cognitive Assessment for Robotic Surgery (ICARS) is the first objective nontechnical skill assessment tool designed specifically for robotic surgery.[22] Using a Delphi consensus of expert robotic and laparoscopic surgeons, ICARS examines 28 nontechnical components that span 4 domains: checklist and equipment, interpersonal skills, cognitive skills, and resource skills. In addition, a unique feature of ICARS is the ability to assess nontechnical performance at both the bedside and surgeon console. This tool has demonstrated a strong correlation with NOTSS as well as its application in surgical crisis simulation. Moreover, Raison et al. provide evidence supporting the ability of ICARS to differentiate amongst novice, intermediate, and expert performance.[22] With the acquisition of additional validity evidence, including the impact of scores on clinical outcomes, this instrument could be integrated into both formative and summative assessments of RUS competency.

Electroencephalography

Recently, electroencephalography (EEG) has been used as a real-time, objective measurement of brain activity during RUS. EEG data is analyzed and categorized into three main cognitive metrics: cognitive state, cognitive workload, and high-level engagement/distraction. Guru et al. found that although different levels of mental faculties are employed depending on the surgical task, higher mental resource usage generally correlates with improved subjective performance on the NASA-TLX.[23] Using the high-level engagement and cognitive state metrics, Guru et al. demonstrated the discriminatory ability of EEG between beginner, intermediate, and expert robotic surgeons completing suturing and knot tying tasks.[24] The same parameters could also distinguish between intermediate and experts when performing urethrovesical anastomosis. Hussein et al. expanded the use of EEG to the surgical mentor and found that when the mentor was not satisfied with the trainee's performance, there was increased EEG activity, suggesting more mental engagement.[25]

DISCUSSION

This review highlights the underlying validity evidence supporting nontechnical skill assessment tools, to facilitate their implementation into a structured RUS curriculum. As demonstrated by this review, nontechnical skills in RUS remain understudied compared to their technical skill counterparts, despite making up a critical component of a surgeon's performance and therefore clinical outcomes. It is important to recognize that RUS creates a unique surgical ecosystem. These studies represent the initial efforts to transition from subjective to objective assessments of cognitive metrics and behavior, with the majority of the validity evidence centered around the broadly implementable NASA-TLX score. We believe that this field will likely continue to flourish as robotics gains more popularity within the surgical community.

Several gaps in the literature have been identified in this review. Firstly, the bedside assistant and console surgeon present their own set of challenges and as such,

TABLE. Validity Evidence of Assessments of Nontechnical Skills

Assessment Tool	Settings Used	Domains Tested	No. of Studies Supporting Validity Evidence	No. of high quality studies (MERSQI ≥ 14)
NASA-TLX	VR-based simulator, dry lab, animal model, live surgery	Mental demand, physical demand, temporal demand, performance, effort, frustration	Content: 9 Response process: 2 Internal structure: 0 Relationship to other variables: 8 Consequence of testing: associated with inattention blindness in peripheral vision	1
Electroencephalography	Dry lab, live surgery	Cognitive engagement, mental workload, mental state	Content: 3 Response process: 0 Internal structure: 0 Relationship to other variables: 3	0
ICARS	Dry lab	WHO checklist completion, console setup, communication and team skills, leadership, decision-making, situational awareness, stress and distractors	Content: 1 Response process: 1 Internal structure: 1 Relationship to other variables: 1	0
NOTSS	Dry lab	Situational awareness, decision making, communication and teamwork, leadership	Content: 0 Response process: 0 Internal structure: 0 Relationship to other variables: 1	0

require distinct nontechnical skill evaluations.[18] ICARS represents a shift toward this direction, however more supporting evidence is needed. The feasibility of conducting assessments of both the console and assisting surgeon during live surgery is also poorly understood at this time. Secondly, although there is an emerging trend in the development of procedure-specific assessment tools for technical skill,[26–28] none exist for nontechnical skill, and this may limit the direct applicability of feedback generated by the assessments. Thirdly, a clear link between cognitive workload and procedural proficiency needs to be established. This will facilitate the characterization of appropriate benchmarks for achieving competency at each stage of RUS training. At present, most studies compare opposite ends of the spectrum of robotic experience. Similar to the landscape of technical skill assessment tools, the next frontier is to develop scoring algorithms sensitive enough to identify trainees (of similar case experience and procedural training) that require remediation. Lastly, the management of intraoperative complications during RUS needs to be further explored. Zattoni et al. provides initial evidence supporting the use of open conversion simulations during robot-assisted radical prostatectomy and highlights the importance of nontechnical skills in minimizing errors and adverse outcomes.[29] Such simulation can be expanded into surgical crises currently being investigated in laparoscopic surgery, including renal hilar vessel and inferior vena cava injuries.[30,31]

Several groups have proposed curricula that address both technical and nontechnical skill acquisition in urology.[32–34] Their approaches are divided into early, intermediate, and advanced phases of training. As more assessment tools are validated, a similar framework should be implemented for RUS. This will range from understanding the robotic surgical milieu and syntax in the early stages to developing leadership as the console surgeon and managing surgical crises in the advanced stages. Only through a structured and deliberate approach to RUS training can we ensure that graduating trainees have reached a minimum level of competency that translates to satisfactory surgical outcomes.

CONCLUSIONS

This review has highlighted a pressing need for investigating assessments of nontechnical skills in RUS. Additional work is needed to better characterize the RUS training continuum. Nontechnical skill assessment tools specific to the robotic surgical ecosystem must be developed and studied across various contexts of surgical training. The impact of these metrics on patient outcomes, operating room efficiency, and even healthcare costs must be evaluated. These tools must also be accessible and universally applicable across RUS training programs. Nontechnical performance should be assessed in conjunction with technical ability as a comprehensive measure of robotic skill acquisition, as they may be independent contributors to end outcomes. Future work must account for both the technical and nontechnical aspects of trainees and surgeons.

CONFLICT OF INTEREST

None

APPENDIX A. SUMMARY OF INCLUDED STUDIES ASSESSING NONTECHNICAL SKILLS

Study	Trainees	Setting of Assessment	Raters	Measurement Tool	MERSQI
Sethi et al. (2009)	15 novice (MS, R) 5 expert (F, S)	Simulator	Self-rate	NASA-TLX	10
Klein et al. (2009)	15 novice	Dry	Self-rate	NASA-TLX, Multiple resources questionnaire	9.5
Hughes-Hallett et al. (2015)	73 S	OR (partial nephrectomy)	Computer-based Self-rate	Inattention blindness, cognition, NASA-TLX	14
Guru et al. (2015)	2 novice 5 intermediate 3 expert	Dry	EEG	Cognitive engagement, mental workload, mental state	11.5
Chowriappa et al. (2015)	22 R 30 F	Simulator Dry	Self-rate	NASA-TLX	13.5
Guru et al. (2015)	1 expert	OR (lysis of adhesions, extended lymph node dissection, urethrovesical anastomosis)	EEG Self-rate	Cognitive engagement, mental workload, mental state, NASA-TLX	12
Zattoni et al. (2015)	2 S	OR simulation (open conversion)	Expert surgeon	Robot movement errors, space conflict, communication errors, lack of leadership, lack of task sequence, loss of sterility, accidental fall of surgical devices	9.5
Mouraviev et al. (2016)	21 R	Simulator Wet (porcine)	Self-rate	NASA-TLX	12
Hussein et al. (2016)	20 S	OR (extended lymph node dissection, urethrovesical anastomosis)	EEG Self-rate	NASA-TLX, mental state, distraction, mental workload	12.5
Lee et al. (2017)	32 T (R, F)	Simulator Dry	Self-rate	NASA-TLX, Multiple resources questionnaire, Short stress state questionnaire	12.5
Raison et al. (2017)	59 novice 6 intermediate 8 expert	OR simulation	Expert surgeon NTS expert	Interpersonal and cognitive assessment for robotic surgery, Nontechnical skills for surgeons	12.5
Sessa et al. (2017)	12 novice 9 expert	Simulator	Self-rate	NASA-TLX	9.5

MS, medical student; R, resident; F, fellow; S, staff; T, trainee.

REFERENCES

1. Zorn KC, Gautam G, Shalhav AL, et al. Training, credentialing, proctoring and medicolegal risks of robotic urological surgery: recommendations of the Society of Urologic Robotic Surgeons. *J Urol*. 2009;182(3):1126–1132. https://doi.org/10.1016/j.juro.2009.05.042.

2. Lee JY, Mucksavage P, Sundaram CP, McDougall EM. Best practices for robotic surgery training and credentialing. *J Urol*. 2011;185(4):1191–1197. https://doi.org/10.1016/j.juro.2010.11.067.

3. Goldenberg MG, Lee JY, Kwong JC, Grantcharov TP, Costello A. Implementing assessments of robotic-assisted technical skill in urologic education: a systematic review and synthesis of the validity evidence. *BJU Int*. 2018. https://doi.org/10.1111/bju.14219.

4. Rogers SO, Gawande AA, Kwaan M, et al. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery*. 2006;140(1):25–33. https://doi.org/10.1016/j.surg.2006.01.008.

5. Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. *World J Surg*. 2008;32(4):548–556. https://doi.org/10.1007/s00268-007-9320-z.

6. Yule S, Parker SH, Wilkinson J, et al. Coaching non-technical skills improves surgical residents' performance in a simulated operating room. *J Surg Educ*. 2015;72(6):1124–1130. https://doi.org/10.1016/j.jsurg.2015.06.012.

7. Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Qual Saf Health Care*. 2009;18(2):104–108. https://doi.org/10.1136/qshc.2007.024760.

8. Russ S, Hull L, Rout S, Vincent C, Darzi A, Sevdalis N. Observational teamwork assessment for surgery. *Ann Surg*. 2012;255(4):804–809. https://doi.org/10.1097/SLA.0b013e31824a9a02.

9. Yule S, Paterson-Brown S. Surgeons' non-technical skills. *Surg Clin North Am*. 2012;92(1):37–50. https://doi.org/10.1016/j.suc.2011.11.004.

10. Ahmed K, Brunckhorst O. Standardising and structuring of robotic surgery curricula: validation and integration of non-technical skills is required. *BJU Int*. 2014;113(5):687–689. https://doi.org/10.1111/bju.12602.

11. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700. https://doi.org/10.1136/BMJ.B2700.

12. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA, 298; 2007. p. 1002. https://doi.org/10.1001/jama.298.9.1002.

13. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *ETS Res Rep Ser*. 1994;1994(2):I–28. https://doi.org/10.1002/j.2333-8504.1994.tb01618.x.

14. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. http://www.stavelandhfe.com/images/NASA-TLX_paper.pdf. Accessed January 14, 2018.

15. Klein MI, Lio CH, Grant R, Carswell CM, Strup S. A mental workload study on the 2d and 3d viewing conditions of the da Vinci surgical robot. In: In: Proceedings of the Human Factors and Ergonomic Society 53rd Annual Meeting; 2009. https://doi.org/10.1518/107118109X12524443345555.

16. Sethi AS, Peine WJ, Mohammadi Y, Sundaram CP. Validation of a novel virtual reality robotic simulator. http://online.liebertpub.com/doi/pdf/10.1089/end.2008.0250. Accessed January 14, 2018.

17. Mouraviev V, Martina Klein B, Eric Schommer B, et al. Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules. *J Robot Surg* 10. doi:10.1007/s11701-015-0540-1.

18. Sessa L, Perrenot C, Xu S, et al. Face and content validity of XperienceTM Team Trainer: bed-side assistant training simulator for robotic surgery. *Updates Surg*. 1 (123456789). doi:10.1007/s13304-017-0509-x.

19. Chowriappa A, Raza SJ, Fazili A, et al. Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: a multi-institutional randomised controlled trial. *BJU Int*. 2015;115(2):336–345. https://doi.org/10.1111/bju.12704.

20. Lee GI, Lee MR. Can a virtual reality surgical simulation training provide a self-driven and mentor-free skills learning? Investigation of the practical influence of the performance metrics from the virtual reality robotic surgery simulator on the skill learning and associated cognitive workloads. *Surg Endosc*. 2018;32(1):62–72. https://doi.org/10.1007/s00464-017-5634-6.

21. Hughes-Hallett A, Mayer EK, Marcus HJ, et al. Inattention blindness in surgery. *Surg Endosc*. 2015;29 (11):3184–3189. https://doi.org/10.1007/s00464-014-4051-3.

22. Raison N, Thomas Wood B, Oliver Brunckhorst B, et al. Development and validation of a tool for non-technical skills evaluation in robotic surgery—the ICARS system. *Surg Endosc*. 31. https://doi.org/10.1007/s00464-017-5622-x.

23. Guru KA, Shafiei SB, Khan A, Hussein AA, Sharif M, Esfahani ET. Understanding cognitive performance during robot-assisted surgery. *Urology*. 2015;86(4):751–757. https://doi.org/10.1016/j.urology.2015.07.028.

24. Guru KA, Esfahani ET, Raza SJ, et al. Cognitive skills assessment during robot-assisted surgery: separating the wheat from the chaff. *BJU Int*. 2015;115 (1):166–174. https://doi.org/10.1111/bju.12657.

25. Hussein AA, Shafiei SB, Sharif M, et al. Technical mentorship during robot-assisted surgery: a cognitive analysis. *BJU Int*. 2016;118(3):429–436. https://doi.org/10.1111/bju.13445.

26. Raza SJ, Field E, Jay C, et al. Surgical competency for urethrovesical anastomosis during robot-assisted radical prostatectomy: development and validation of the robotic anastomosis competency evaluation. *Urology*. 2015;85(1):27–32. https://doi.org/10.1016/j.urology.2014.09.017.

27. Lovegrove C, Novara G, Mottrie A, et al. Structured and modular training pathway for robot-assisted radical prostatectomy (RARP): validation of the RARP assessment score and learning curve assessment. *Eur Urol*. 2016;69(3):526–535. https://doi.org/10.1016/j.eururo.2015.10.048.

28. Hussein AA, Ghani KR, Peabody J, et al. Development and validation of an objective scoring tool for robot-assisted radical prostatectomy: prostatectomy assessment and competency evaluation. *J Urol*. 2017;197(5):1237–1244. https://doi.org/10.1016/j.juro.2016.11.100.

29. Zattoni F, Guttilla A, Crestani A, et al. The value of open conversion simulations during robot-assisted radical prostatectomy: implications for robotic training curricula. *J Endourol*. 2015;29(11):1282–1288. https://doi.org/10.1089/end.2015.0435.

30. Lee JY, Mucksavage P, McDougall EM. Simulating laparoscopic renal hilar vessel injuries: preliminary evaluation of a novel surgical training model for residents. *J Endourol*. 2012;26(4):393–397. https://doi.org/10.1089/end.2011.0432.

31. Goldenberg MG, Fok KH, Ordon M, Pace KT, Lee JY. Simulation-based laparoscopic surgery crisis resource management training—predicting technical and nontechnical skills. *J Surg Educ*. 2017. https://doi.org/10.1016/J.JSURG.2017.11.011.

32. Arora S, Lamb B, Undre S, Kneebone R, Darzi A, Sevdalis N. Framework for incorporating simulation into urology training. *BJU Int*. 2011;107(5):806–810. https://doi.org/10.1111/j.1464-410X.2010.09563.x.

33. Rashid P, Gianduzzo TRJ. Urology technical and non-technical skills development: the emerging role of simulation. *BJU Int*. 2016;117:9–16. https://doi.org/10.1111/bju.13259.

34. Aydin A, Raison N, Khan MS, Dasgupta P, Ahmed K. Simulation-based training and assessment in urological surgery. *Nat Rev Urol*. 2016;13(9):503–519. https://doi.org/10.1038/nrurol.2016.147.