Implementing Assessments of Robotic-Assisted Technical Skill in Urologic Education: A Systematic Review and Synthesis of the Validity Evidence

Mitchell G Goldenberg[1], Jason Y Lee[1], Jethro CC Kwong[2], Teodor P Grantcharov[3], Anthony Costello[4]

[1]*Division of Urology, Department of Surgery, University of Toronto, Toronto, Canada*

[2]*Faculty of Medicine, University of Toronto, Toronto, Canada*

[3]*Division of General Surgery, Department of Surgery, University of Toronto, Toronto, Canada*

[4]*Department of Surgery/Urology, University of Melbourne, Royal Melbourne Hospital, Melbourne, Australia*

Correspondence: Mitchell Goldenberg, MBBS
St. Michael's Hospital
30 Bond Street, 16CC-056
Toronto, Ontario
Canada M5B 1W8
Tel: 416 864-5748
Fax: 416 864-5343

Abstract: 180
Word Count: 4,617

DR. MITCHELL G. GOLDENBERG (Orcid ID : 0000-0002-4601-5721)

Article type : Surgical Education

Correspoding author mail id:- mitchg.11@gmail.com

Article Category: Professional Innovation

Abstract

Objectives: To systematically review and synthesize the validity evidence supporting intraoperative and simulation-based assessments of technical skill in urologic robotic-assisted surgery (RAS), and make evidence-based recommendations for the implementation of these assessments in urologic training.

Materials and Methods: A literature search of the MEDLINE, PsycINFO and Embase databases was performed. Articles using technical skill and simulation-based assessments in RAS were abstracted. Only studies involving urology trainees or faculty were included in the final analysis.

Results: Multiple tools for the assessment of technical robotic skill have been published, with mixed sources of validity evidence to support their use. These evaluations have been used in both the ex vivo and in vivo settings. Performance evaluations range from global rating scales to psychometrics, and assessments are carried out through automation, expert analysts, and crowdsourcing.

Conclusion: There have been rapid expansions in approaches to robotic technical skills assessment, both in simulated and clinical settings. Alternative approaches to assessment in RAS such as crowdsourcing and psychometrics remain under investigation. Evidence to support the use of these metrics in high-stakes decisions is likely insufficient at present.

Keywords: Urology training; Robotic Surgery; Technical Skill Assessment; Simulation; Education; Competency

1. Introduction

Surgical education is experiencing a huge shift from Halstead's apprenticeship model introduced over 100 years ago to the current climate of competency-based education. A trainee must exhibit clinical competence, and in surgical education this includes both the technical and non-technical skills needed to safely carry out any number of procedures. Evidence linking technical performance to patient outcomes and safety has drawn the public's attention, reflected by recent efforts to allow patient access to video footage of surgical procedures[1]. These developments have significantly altered the way we in which approach research in surgical assessment and curriculum design.

More than in any other surgical field, robotic-assisted surgery (RAS) has been rapidly embraced by the urologic community. It is quickly becoming the most common approach to many operations, including prostatectomy, partial nephrectomy, pyeloplasty, cystectomy, and retroperitoneal node dissection (RPLND)[2]. Its predominant use continues to be for prostate cancer, where Robotic-Assisted Radical Prostatectomy (RARP) has become the gold standard in the surgical management of localized prostate cancer in most of the developed world[2]. The dynamic growth of this surgical technique has had a wide impact on practicing urologists and surgical residency and fellowship programs alike. The need for formalized RAS training has also resulted in increased need for assessments of skill, both formative and summative. Despite the continued creation of new tools to assess performance, important questions remain unanswered; how do we effectively incorporate RAS training programs into urology residency curricula? How do we appropriately credential practicing urologists wishing to perform robotic surgery?

How do we incorporate the most effective education programs in the urology residency curricula? Most recent Urology residency graduates will not have had an immersive experience in robotic surgery. Those Urologists who passed their Board or Fellowship exams 10 years ago have had to acquire the required robotic skills in a very unstructured transitioning surgical landscape. It may even be appropriate to include robotic surgical education curricula late in medical school training. This would permit early recognition of those students with aptitude in surgery to be identified using the metrics outlined in this manuscript.

For an objective assessment of robotic skill to be applicable in training, privileging or accreditation, it is essential to build a 'validity argument' supporting its use. Messick's Conceptual Framework is an acceptable way to construct such an argument, through the assembly of various sources of validity evidence, specifically content, response process, internal consistency, relationship to other variables, and consequences[3]. This type of framework replaces the now outdated Cronbach Taxonomy of validity (predictive, concurrent, content and construct validity), by seeing validity as a dynamic or fluid concept that must be argued in different assessment environments.

Like any procedural assessment rubric, the tools used to evaluate robotic skill employ a combination of global rating scales (GRS) and task-specific checklists to assess trainee competencies[4,5]. Using trained expert analysts, GRS can be superior in both accuracy and reliability across a wide variety of procedure-types when compared to checklists[6]. Despite this the validity evidence supporting objective assessments of technical skill remains insufficient to warrant their use in high-stakes decisions such as progression through competency-based training or credentialing[7]. It is vital to create a validity argument in support of these approaches when considering their inclusion in summative assessments in training and beyond[8].

While both technical and non-technical skills are essential in the training of future robotic surgeons[9],this article focuses on technical skill assessments only. The objective of this article is to provide a focused review of the available tools for assessment of robotic surgical technical skill currently available to surgeon educators, and to critically appraise the supporting literature to determine how best to implement these assessment tools into residency and fellowship curricula.

2. Methods

Eligibility criteria

Articles assessing the robotic surgical skill of urologic trainees (medical students, residents, fellows) or faculty urologists were included. Studies assessing robotic skills in other surgical specialties that did not include urology participants were excluded. Studies primarily assessing non-technical skills were excluded from this review, although the search was designed to capture these studies for future work. Studies published in peer-reviewed journals were included in the analysis, and unpublished abstracts were included only if it was determined that they contained data contributing to the validity of the assessment being studied. Randomized control trials and observational studies, including cohort, case–control, case series and cross-sectional studies, were all eligible for inclusion.

Information sources

One author conducted a search in Ovid MEDLINE, Embase Classic, PsycINFO and the Cochrane Library. The search was carried out on July 18th, 2017.

Search

Medical subject headings (MeSH) terms used in the search included ‘communication’, ‘clinical competence’, ‘curriculum’, ‘education, medical’, ‘surgical procedures’, ‘education, medical, graduate’, ‘educational measurement’, ‘medical errors’, ‘nephrectomy’, ‘patient simulation’, ‘prostatectomy’, ‘robotic’, ‘robotic surgery’, ‘robotic surgical procedures’, ‘robotics’, ‘skill’, ‘surgery’, ‘non-technical skill’, ‘cognitive skill’, ‘technical’, ‘technical skill’, ‘urologists’, ‘urology’. Titles of articles resulting from the search and corresponding abstracts were reviewed initially and articles eligible for full-text review were identified. These articles were then analyzed further to ensure that no articles referenced therein were missed for inclusion in the full-text review. Duplicates were identified and removed.

Study selection

Any study in the medical or surgical literature that assessed the robotic surgical skill of urologic trainees or faculty, involving original research and described in English, were included. Opinion letters, editorials, case reports, reviews, and letters to the editor were excluded. References used in previous review articles were assessed and those that met the inclusion criteria were incorporated in the analysis. Articles that looked at outcomes only were also excluded. Two authors considered the articles for inclusion independently, and any disagreements were resolved by consensus.

Data collection process

Data were abstracted from the included studies systematically, including sample size, participants, assessment used, study setting, rater information, and assessment design and implementation relevant to various sources of validity evidence.

Quality assessment

The Medical Education Research Study Quality Instrument (MERSQI) was used to assess the quality of the included articles[10]. The MERSQI scores quality over eight domains: study design, institutions sampled, response rate, type of data, validity evidence for evaluation of instrument scores, sophistication of data analysis, appropriateness of data analysis, and assessment outcome.

Validity Evidence

We used Messick's validity framework[3] to structure the evidence supporting the application of these assessment tools in robotic surgery. These sources of test validity include content, response process, internal structure, relationships to other variables, and consequences of testing. Use of this framework allowed us to put forward our own, evidence-guided recommendations on how best to implement these assessments into formal training curricula.

3. Results

Our initial search yielded 566 articles. After two independent authors reviewed titles and abstracts, 282 articles were selected for full review to determine inclusion status. Following full text review and cross-checking of article references, 85 studies were included in the final analysis (Figure-1). The included articles are displayed in Appendix-1, subdivided into assessments of technical skill and computer-based virtual reality (VR) assessments.

3.1 Technical Skill Assessments

Table-1.1 summarizes the validity evidence supporting the seven non-time-based technical skill assessment tools used in urological robotic surgery. The Global Evaluative Assessment of Robotic Skills (GEARS) tool, developed by Goh et al, has been applied to urological assessments on multiple occasions[4,11-29], and has the strongest validity argument supporting its use in the assessment of robotic skill. Its generic framework has allowed it to become a widely accepted method of assessment across multiple procedures and even across specialties[15,30,31]. Notably, evidence supports its ability to discriminate amongst staff surgeons of differing case volume[16], as well as across a single surgeons learning curve[13]. The vast majority of literature using the GEARS score has found it to be a reliable assessment method[4,16,24,25,27,28,32]. However, a study of robotic renal hilar dissection using oriented expert raters showing poor internal consistency[17], and Hung and colleagues found that while trainee self-assessments and faculty evaluations correlated weakly, inter-faculty reliability was better when assessing residents (ICC=0.77) and fellows (ICC=0.45)[21]. As shown in Table-1.1, it is the only technical skill assessment tool that has supporting consequences evidence, having been used to both predict clinical outcomes in a retrospective case-control study, and impact residency match-rankings when applied to a cohort of medical students. The Global Operative Assessment of Laparoscopic Skills (GOALS) [11,33-35], a laparoscopic-specific GRS that served as the underlying model for the GEARS, was also used in robotic skills assessment by Hung et al[34], with the addition of two robotic-specific domains, instrument awareness and precision and camera awareness and precision. Their randomized control trial demonstrated that baseline performance on a virtual reality simulation scenario correlated with performance on a porcine model. Tunitsky et al[35] demonstrated GOALS ability to discriminate between procedural expert surgeons and robotic expert surgeons performing a simulated robotic ureteral anastomosis, providing evidence that this GRS may be able to adequately evaluate procedural-specific constructs. The Objective Structured Assessment of Technical Skill (OSATS) tool[11,18,36-41], originally developed at the University of Toronto for a 'bench-station' examination of basic surgical skills[42], has been used to assess robotic technical skill, with multiple studies providing various types of validity evidence, across simulation, laboratory, and clinical environments. Siddiqui and colleagues[5] added robotic-specific metrics to the OSATS tool, using 5 dry-lab 'drills' to assess robotic skill across 4 domains, terming their modification 'R-OSATS'. They demonstrated its relationship to other variables by comparing scores to training level and console experience. Their tool also exhibited excellent inter-rater reliability (Cronbach's $\alpha = 0.91$). A RARP-specific assessment tool, the Robotic Anastomosis Competency Evaluation (RACE)[16,43] was developed by Raza et al, and uses global ratings across 5 domains to assess specific skills needed to complete the vesicourethral anastomosis step of the RARP. While their tool could discriminate between trainees of different experience, the reliability of their tool was only moderate ($\alpha = 0.62$). The RARP Assessment Score[44] was developed by an international group using the Healthcare Failure Mode Effect Analysis (HFMEA). The HFMEA[45] is a method of human risk analysis, which allowed the authors to identify high-risk steps of the procedure to include in their assessment of trainees taking part in a European robotics fellowship. However, the small numbers of participants in their study makes interpretation of their data difficult at this stage. The Prostatectomy Assessment Competency Evaluation (PACE) is the product of a Delphi consensus of international urologic oncologists[46]. Like the RARP Assessment Score, this tool is procedure specific. Each step of the procedure is rated using a 5-point Likert scale, with agreed upon anchor points for scores of 1, 3 and 5. Finally, the Assessment of Robotic Console Skills (ARCS) was developed in collaboration with Intuitive Surgical as a global rating scale to more-specifically assess console skills, including optimization of field of view and workspace, and basic energy pedal skills[47]. Their initial validation study demonstrated the ARCS ability to discriminate between staff surgeons of less than 100 versus greater than 100 completed robotic-assisted cases.

In addition to these GRS assessments, studies used weighted combinations of time and error[48-60] (similar to the Fundamentals of Laparoscopic Surgery, FLS[61]) and 'end-product' scores[50,58,59,62] to assess technical performance.

3.2 Computer-Based VR Assessment

Table 1.2 outlines the commercially available simulation platforms and scoring metrics for robotic surgery with literature supporting their use in training urologists. The field of robotic simulation is well established, with multiple developers offering platforms to the public, each with its own unique features, strengths and weaknesses[63].

Intuitive Surgical (Sunnyvale, CA), designer of the daVinci System, is responsible for the daVinci Surgical Simulator (dVSS)[14,15,20,22,29,34,37,53,64-81]. This robotic simulator fits directly onto the surgeon console, allowing the trainee to sit at the same controls he or she would be using in the operating room. It has the disadvantage of not being available if the console is being used in the operating room, as it cannot be used independently of the console[82]. The dVSS is the result of collaboration. The software used by the dVSS was developed initially in conjunction with the Mimic group, and so many similarities are found between these platforms in terms of metrics assessed and the user interface (UI). In 2009, Lerner and colleagues[83] showed that a cohort trained on the dV-Trainer® performed similarly to those trained on the dVSS, and they achieved similar results on dry-lab tasks. This outcome may reflect the similarities in their software design and UI. Additionally, the selection and creation of the tasks used by the dVSS was made in conjunction with the Simbionix group. In a study by Amirian et al[59], the Simbionix suturing module (SSM), running on the dVSS training software, was able to demonstrate improvement from baseline in a group of robotic novices. Lee et al developed a four-week training curriculum, the Basic Skills Training Curriculum (BSTC)[84], which employed the dVSS system to compare a time-based method of assessment with a competency/proficiency-based method in surgeons of various training levels at the University of Toronto. Hung and colleagues[70] used visual analogue scales (VAS) to establish the functional task alignment of the dVSS, and their study showed again that this simulation platform can distinguish between experts and novices. In a subsequent study, this group demonstrated that assessments with the dVSS have clinical consequences[34], by correlating baseline trainee skill with ex vivo tissue performance after the completion of a dVSS dry-lab curriculum.

Another popular robot-specific platform is the dV-Trainer® developed by Mimic (Seattle, WA) [54,73,83,85-93]. Initial validation studies[88,94] provided evidence that the simulator was able discriminate between expert and novice robotic surgeons. In a 2012 study, Lee and colleagues[54] demonstrated that dV-Trainer® performance correlates with actual daVinci console performance at dry-lab tasks. New initiatives from Mimic include the Xperience Team Trainer, which includes an assistant laparoscopic simulator that integrates a communication element into the simulation experience.[33]

Simbionix (Israel) has developed multiple procedural simulators across different specialties, including the RobotiX Mentor Platform®. Like the dV-Trainer®, it too is a stand-alone platform and can incorporate a laparoscopic assistant simulator. Validity evidence for its use comes from a study from Whittaker and colleagues[95], in which they were able to demonstrate significant score differences between novices and experts, using two simulated modules and employing domains of assessment from the Foundations of Robotic Surgery curriculum (FRS). Simbionix-developed software that allows trainees to complete virtual reality steps of the radical prostatectomy have been recently integrated into both the RobotiX and dVSS platforms.

The Robotic Surgery Simulator (RoSS) [96,97], made by Simulated Surgical Systems (San Jose, CA), is another simulator, and unlike the dVSS, it is a standalone platform.

While it is not identical to the daVinci console used by the dVSS, it is modeled after it, and subsequently has similar task alignment[97]. It was developed with the Roswell Park group in Buffalo, NY, and this group has demonstrated that the RoSS has the ability to predict performance on another simulator[98], as well as intraoperative ability[99]. Finally, the RoSS simulator has now integrated the RSA-score assessment tool[96], developed through the FSRS group as described above, further adding to its applicability to robotic curricula.

The final platform designed specifically for robotic surgery simulation is the Sim surgery Education Platform (SEP) Robot Simulator (Oslo, Norway). This is a less utilized platform, and the evidence for it has been mixed[100-102]. Studies have been able to show that novices performed consistently poorer when compared with a cohort of experts on the SEP platform.

A unique example of laparoscopic simulator technology being applied to robotic surgery is the ProMIS system[103-105]: a platform that measures efficiency of task completion such as total distance of instrument arm movements and smoothness of motion[103]. A urology-specific example of its use in robotics comes from a study by Jonsson et al[104], who's group showed that the ProMIS simulator was able to discriminate between novices and experts at a dry-lab vesicourethral anastomosis model. This article further added to its validity evidence by comparing the smoothness of motion metric between groups, to the more conventional measurement of time to task completion.

Key differences exist between these simulators. A unique and important property of computer-based VR simulators is the ability to automatically track instrument movements. The dV-Trainer® and SEP simulators measure the force with which the instruments are used, as well as instrument collisions, an important issue with robotic surgery where haptic feedback does not exist. The dVSS contains the 'system settings' and 'wrist manipulation' measurements, performance domains specific to RAS. Interesting assessments incorporated into the SEP platform are tightening and winding stretch. These measure the amount of tension used in knot tying, an important and advanced robotic skill. Finally, the Mscore assessment rubric developed by Mimic and incorporated into the dV-Trainer (older versions of Mscore also found on the dVSS) allow surgeon mentors and educators the ability to individualize training curricula with development of customized tasks and modular learning activities and deliberate practice sessions based on trainee needs.

3.3 Novel Assessment Methodologies

Novel methods of assessing robotic surgical skill have been introduced in the recent literature. We describe four such innovations here, and they are summarized in Table-2.

Crowdsourcing

An exciting but controversial area of assessment being established in robotic technical skill assessment is 'crowdsourcing'[106]. This method uses members of the public, medically trained or not, to make judgments on surgical skill and technique. Consistently, studies have shown that these groups of people, often referred to as 'turkers', have not only excellent internal consistency, but also have ratings correlative to those of expert surgeons[106]. C-SATS[28], an online platform that utilizes this method, has been used in multiple surgical fields, including laparoscopy and robotics. Recently, efforts from the Michigan Urological Surgery Improvement Collaborative (MUSIC) have applied this method of assessment to robotic radical prostatectomy[16], showing that crowdsourcing is applicable to assessment of this procedure using GEARS. However, it was noted that the 'crowd' was less willing to rate participants as either very poor or very good performers, which was not the case for expert raters. This phenomenon may question the use of this method in summative or high-stakes assessments, where distinguishing between high and low performers is imperative. Additionally, there is a considerable cultural barrier to overcome in this case, as experienced surgeons may doubt the ability of non-medically trained crowd workers to potentially judge whether surgeons are competent at performing advanced surgical procedures. Certainly, there will be more investigation into this assessment modality, including whether crowd-derived judgments can reliably predict not only expert opinion but also patient outcomes.

Machine Learning

A study by Kumar et al[107] used a form of artificial intelligence (AI), Support Vector Machines (SVM), to assess the robotic workspace adjustment and camera manipulation of trainees performing a variety of tasks on the robotic console. They found that their algorithm had a classification accuracy of over 95% for workspace adjustment, and over 88% for camera manipulation. Despite some study limitations, the use of AI in skill assessments is a rapidly growing and promising field of research.

Motion/Contact Vibrations

Many groups across all surgical platforms are looking for methods of assessment that use purely objective psychometrics to eliminate the inherent bias of human judges. In our review, Gomez and colleagues[18] had some success using contact vibration as a surrogate for robotic skill in a series of dry-lab tasks. Their study demonstrated that lower vibration and force-derived metrics were recorded in their cohort of experienced robotic surgeons as compared to novices. This novel evaluation method showed good construct validity in 10 out of 15 metric-task correlations, demonstrating that this purely objective method has utility in formative skill assessments. However, this and similar unidimensional psychomotor assessments may not reflect the full competence, or lack thereof, and must demonstrate correlation with patient outcomes before they are accepted on the main stage of surgical assessment.

Armrest Load

Two studies from Yang et al.[73,108] quantified armrest load and surgeon ergonomics as methods of both assessment and educational intervention in robotic surgery training. They found they could distinguish between surgeons with different robotic experience in a simulated environment, as well as shorten the simulation-based learning curve of novice trainees by building in a real-time feedback mechanism that alerts the trainee about excessive weight applied to the console armrest. This metric has potential as a means of both improving trainee acquisition of technical competency and complementing assessments of surgeon skill in training curricula.

3.4 Literature Quality Assessment

The mean MERSQI score for all included articles was 12.8, which falls short of the 14/18 mark that indicates ‘high quality’. Articles found to have a score of 14 or higher are detailed in Table-3.

4. Discussion

This review has highlighted the various assessment methods that exist in evaluating technical skill when performing robotic surgery in urology. This area of research is still actively evolving, and while this article has summarized the methods used to date, we expect that applications and diversity of these instruments will continue to expand and develop as the paradigm of competency-based training becomes the standard.

We have outlined the various efforts made in assessing technical skill in urologic robotic surgery, and while the literature is diverse, we have shown some homogeneity in the underlying principles of assessment being employed. As in most studies assessing technical skill, global rating scales continue to be more popular than task-specific checklists, due to their broader applicability and ease of use[6].

Although many of these assessment tools can be applied across all types of robotic surgery, urology will likely lead the movement toward the use of these assessments in surgeon accreditation, as opposed to its current place in the formative setting only. Educators and licensing stakeholders will pay attention in urology especially, as the role of surgeon performance in patient safety and outcomes continues to be investigated in this space[13]. This emerging evidence will likely lead to the incorporation of assessments of technical and non-technical skill into licensing practices at a local or national level[109]. As of now, the accreditation process remains under the sole control of the hospitals[2], and there is no established use of summative technical skill assessments in robotic surgery for the purposes of credentialing.

There are specific limitations of this review and the included research presented. A major issue that is prevalent throughout the robotic assessment literature is the comparison of novice and expert surgeons as a source of validity evidence. In order to frame an assessment in a specific context, i.e. low-stakes vs. high-stakes, it is crucial that the assessment construct be clearly defined. Making decisions of competency within a training program requires the chosen assessment to distinguish between trainees who have met a predefined set of criteria from those that require further remediation. In contrast, an assessment designed for credentialing robotic surgeons after training must be able to distinguish between those who will have satisfactory patient safety and clinical outcomes and those that do not. Unfortunately, much of the literature choses to compare groups at the extremes of skill to allow for highly statistically significant differences in 'scores' between cohorts. Secondly, it is important to note that the internal structure and response process validity for simulators is often hard to quantify. Although computer-generated and algorithm-based scoring metrics are assumed to be accurate and reliable, it is still essential that manufacturers and academics strive to provide this validity evidence as robustly as possible, by clearly describing how their scoring components are tabulated and weighted, and any quality control process that are undertaken in the development of scoring algorithms.

Importantly, most studies in this review contribute at least one source of validity evidence for their described assessment tool, as shown in Table-1. However, gaps in the supporting evidence are present in the majority of these studies, and emphasis should be placed moving forward on addressing this. Despite all studies contributing one or more source of validity evidence for a given assessment, many various data elements that make up each of Messick's five domains of validity were vastly underrepresented.[110] Of note, internal structure and response process evidence was fairly homogenous in nature across the included literature. While interrater reliability statistics were more commonly reported, other important internal structure data such as internal consistency (reliability across the domains of the assessment tool) and test-retest reliability (reliability across different sittings or versions of the assessment) were rarely included or described in these studies. Additionally, crucial components of response process evidence such as rater data analysis (understanding rater disagreements or inconsistencies) and effects of rater training (comparison of scores between trained and untrained raters) were also not addressed by most of these studies. Typically, response process evidence in these studies consisted only of descriptions of rater training, and the use of video capture to ensure quality control of testing data. These gaps in evidence may reflect the investigator's use of outdated taxonomies of validity when designing these studies, including decisions around the type of data to calculate and report in their manuscript.

4.1 Recommendations

Using Messick's Conceptual Framework of Validity[3], we have systematically gathered and quantified the validity evidence supporting technical and computer-based VR assessments of robotic surgical skill, to provide evidence-based recommendations on how best to implement these assessment tools in postgraduate training and, in future, credentialing practices.

It is clear from our review that assessments of technical skill using the GEARS metric are strongly supported with robust validity evidence in a wide range of settings, from ranking medical students in the residency match to distinguishing 'high' and 'low' performances of a single, high-volume surgeon. It provides reliable ratings of trainee or faculty performance in real-time assessments in the lab or operating room, or when used in video-based evaluation by expert raters or laypeople through crowdsourcing. However, it is important to note that while many studies report a high to very-high interrater reliability, this is not true of all the included literature. We must stress to educators the importance of training faculty in the use of these assessment rubrics, and early identification of raters who are outliers in their scoring of trainee technical skill. Another option for technical skill evaluation is the OSATS tool, long seen as a gold-standard amongst GRS assessments. This scale has been used in multiple settings in the literature, and has an excellent evidence-base when applied in all testing environments, including dry lab, simulation/VR, and the operating room. Its broadly applicable domains allow it to be used and easily compared with assessments in open and laparoscopic surgery, making it an attractive option for evaluating technical competency across multiple surgical platforms.

It is difficult to provide a single recommendation on computer-based VR assessment, but the validity evidence for both the dVSS and the dV-Trainer systems in low-stakes assessments is strong. Both platforms have been shown to distinguish between trainees and surgeons of differing skill levels, and both have demonstrated response process validity through test-retest methodology and correlation of computer-generated scores with human ratings. Like the GEARS score, these platforms can be used in the training and assessement of participants with a range of robotic surgery experience, but most of the literature supports use in postgraduate education rather than in high-stakes assessments, such as credentialing, as evidence of their ability to predict clinical outcomes is currently lacking.

5. Conclusion

As the competency-based education model of surgical training continues to become more universal[111-113], it is imperative that educators understand not only the milestones set forth by their governing bodies, but also the methods in which these milestones are defined. We have provided a summary of the current literature describing technical skill assessments in urological robotic surgery, and provide evidence-based recommendations of how one may implement these into a competency-based curriculum. Competency in surgical skill must be defined by content experts, through objective means, and the validity evidence of the assessment tools discussed here should give educational stake-holders confidence in making judgments on their trainee’s ability. Despite this, the question of how to best create summative assessments of surgical skill remains unanswered. As demonstrated in this review, there are efforts on multiple fronts, from the simulation lab to the operating room.

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

Conflicts of Interest: Anthony Costello advises Device Technology Australia on an advisory board; Teodor Grantcharov reports grants from Ethicon Canada, Medtronic Canada, Olympus Canada, Takeda Canada and Intuitive Surgical, outside the submitted work; In addition, Dr. Grantcharov has a patent PCT/CA2015/000504 pending, and is the Founder and equity holder in SST Inc. There are no other conflicts to declare.

References

1. Langerman A, Grantcharov TP. Are We Ready for Our Close-up?: Why and How We Must Embrace Video in the OR. Annals of Surgery. March 2017. doi:10.1097/SLA.0000000000002232.
2. 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. JURO. 2009;182(3):1126-1132. doi:10.1016/j.juro.2009.05.042.
3. Messick S. Validity of Psychological Assessment. 1994.
4. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global Evaluative Assessment of Robotic Skills: Validation of a Clinical Assessment Tool to Measure Robotic Surgical Skills. The Journal of Urology. 2012;187(1):247-252.
5. Siddiqui NY, Galloway ML, Geller EJ, et al. Validity and Reliability of the Robotic Objective Structured Assessment of Technical Skills. Obstetrics & Gynecology. 2014;123(6):1193-1199. doi:10.1097/AOG.0000000000000288.
6. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Academic Medicine. 1998;73(9):993.
7. Hatala R, Cook DA, Brydges R, Hawkins R. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. Advances in Health Sciences Education. 2015;20(5):1-27. doi:10.1007/s10459-015-9593-1.

8. 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(3):195-207. doi:10.1023/A:1009849528247.

9. Tiferes J, Hussein AA, Bisantz A, et al. The Loud Surgeon Behind the Console: Understanding Team Activities During Robot-Assisted Surgery. J Surg Educ. 2016;73(3):504-512. doi:10.1016/j.jsurg.2015.12.009.

10. Cook DA, Reed DA. Appraising the Quality of Medical Education Research Methods. Academic Medicine. 2015;90(8):1067-1076. doi:10.1097/ACM.0000000000000786.

11. Vernez SL, Huynh V, Osann K, Okhunov Z, Landman J, Clayman RV. C-SATS: Assessing Surgical Skills Among Urology Residency Applicants. J Endourol. October 2016:end.2016.0569. doi:10.1089/end.2016.0569.

12. Mills JT, Hougen HY, Bitner D, Krupski TL, Schenkman NS. Does Robotic Surgical Simulator Performance Correlate With Surgical Skill? J Surg Educ. June 2017. doi:10.1016/j.jsurg.2017.05.011.

13. Goldenberg MG, Goldenberg L, Grantcharov TP. Surgeon Performance Predicts Early Continence After Robot-Assisted Radical Prostatectomy. J Endourol. 2017;31(9):858-863. doi:10.1089/end.2017.0284.

14. Hung AJ, Jayaratna IS, Teruya K, Desai MM, Gill IS, Goh AC. Comparative assessment of three standardized robotic surgery training methods. BJU Int. 2013;112(6):864-871. doi:10.1111/bju.12045.

15. Ramos P, Montez J, Tripp A, Ng CK, Gill IS, Hung AJ. Face, content, construct and concurrent validity of dry laboratory exercises for robotic training using a global assessment tool. BJU Int. 2014;113(5):836-842. doi:10.1111/bju.12559.

16. Ghani KR, Miller DC, Linsell S, et al. Measuring to Improve: Peer and Crowd-sourced Assessments of Technical Skill with Robot-assisted Radical

Prostatectomy. European Urology. 2016;69(4):547-550. doi:10.1016/j.eururo.2015.11.028.

17. Powers MK, Boonjindasup A, Pinsky M, 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. Journal of Endourology. December 2015:end.2015.0665-end.2015.0666. doi:10.1089/end.2015.0665.

18. Gomez ED, Aggarwal R, McMahan W, Bark K, Kuchenbecker KJ. Objective assessment of robotic surgical skill using instrument contact vibrations. Surg Endosc. July 2015:1-13. doi:10.1007/s00464-015-4346-z.

19. Aghazadeh MA, Jayaratna IS, Hung AJ, et al. External validation of Global Evaluative Assessment of Robotic Skills (GEARS). Surg Endosc. 2015;29(11):3261-3266. doi:10.1007/s00464-015-4070-8.

20. 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(3):475-481. doi:10.1111/bju.13511.

21. Hung AJ, Bottyan T, Clifford TG, et al. Structured learning for robotic surgery utilizing a proficiency score: a pilot study. World J Urol. 2017;35(1):27-34. doi:10.1007/s00345-016-1833-3.

22. Volpe A, Ahmed K, Dasgupta P, et al. Pilot Validation Study of the European Association of Urology Robotic Training Curriculum. European Urology. 2015;68(2):292-299. doi:10.1016/j.eururo.2014.10.025.

23. Hung AJ, Shah SH, Dalag L, Shin D, Gill IS. Development and Validation of a Novel Robotic Procedure Specific Simulation Platform: Partial Nephrectomy. The Journal of Urology. 2015;194(2):520-526. doi:10.1016/j.juro.2015.02.2949.

24. Holst D, Kowalewski TM, White LW, et al. Crowd-Sourced Assessment of Technical Skills: An Adjunct to Urology Resident Surgical Simulation Training. Journal of Endourology. 2015;29(5):604-609. doi:10.1089/end.2014.0616.

25. 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(11):1295-1301. doi:10.1089/end.2015.0191.

26. 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. doi:10.1111/bju.12704.

27. Whitehurst SV, Lockrow EG, Lendvay TS, et al. Comparison of Two Simulation Systems to Support Robotic-Assisted Surgical Training: A Pilot Study (Swine Model). Journal of Minimally Invasive Gynecology. 2015;22(3):483-488. doi:10.1016/j.jmig.2014.12.160.

28. 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. Journal of Endourology. April 2015:150413093359007–6. doi:10.1089/end.2015.0104.

29. Dubin AK, Smith R, Julian D, Tanaka A, Mattingly P. A Comparison of Robotic Simulation Performance on Basic Virtual Reality Skills: Simulator Subjective Versus Objective Assessment Tools. Journal of Minimally Invasive Gynecology. July 2017. doi:10.1016/j.jmig.2017.07.019.

30. Aghazadeh MA, Jayaratna IS, Hung AJ, et al. External validation of Global Evaluative Assessment of Robotic Skills (GEARS). Surg Endosc. 2015;29(11):3261-3266. doi:10.1007/s00464-015-4070-8.

31. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure

robotic surgical skills. The Journal of Urology. 2012;187(1):247-252. doi:10.1016/j.juro.2011.09.032.

32. Vernez SL, Huynh V, Osann K, Okhunov Z, Landman J, Clayman RV. C-SATS: Assessing Surgical Skills Among Urology Residency Applicants. Journal of Endourology. 2017;31(S1):S–95–S–100. doi:10.1089/end.2016.0569.

33. Xu S, Perez M, Perrenot C, Hubert N, Hubert J. Face, content, construct, and concurrent validity of a novel robotic surgery patient-side simulator: the Xperience™ Team Trainer. Surg Endosc. 2016;30(8):3334-3344. doi:10.1007/s00464-015-4607-x.

34. Hung AJ, Patil MB, Zehnder P, et al. Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. The Journal of Urology. 2012;187(2):630-637. doi:10.1016/j.juro.2011.09.154.

35. Tunitsky E, Murphy A, Barber MD, Simmons M, Jelovsek JE. Development and validation of a ureteral anastomosis simulation model for surgical training. Female Pelvic Med Reconstr Surg. 2013;19(6):346-351. doi:10.1097/SPV.0b013e3182a331bf.

36. Vlaovic PD, Sargent ER, Boker JR, et al. Immediate impact of an intensive one-week laparoscopy training program on laparoscopic skills among postgraduate urologists. JSLS. 2008;12(1):1-8.

37. Korets R, Mues AC, Graversen JA, et al. Validating the use of the Mimic dV-trainer for robotic surgery skill acquisition among urology residents. Urology. 2011;78(6):1326-1330. doi:10.1016/j.urology.2011.07.1426.

38. Tarr ME, Rivard C, Petzel AE, et al. Robotic objective structured assessment of technical skills: a randomized multicenter dry laboratory training pilot study. Female Pelvic Med Reconstr Surg. 2014;20(4):228-236. doi:10.1097/SPV.0000000000000067.

39. Phé V, Cattarino S, Parra J, et al. Outcomes of a virtual-reality simulator-training programme on basic surgical skills in robot-assisted laparoscopic surgery. Int J Med Robot. 2017;13(2):e1740. doi:10.1002/rcs.1740.

40. Alemozaffar M, Narayanan R, Percy AA, et al. Validation of a Novel, Tissue-Based Simulator for Robot-Assisted Radical Prostatectomy. Journal of Endourology. 2014;28(8):995-1000. doi:10.1089/end.2014.0041.

41. Rashid HH, Leung Y-YM, Rashid MJ, Oleyourryk G, Valvo JR, Eichel L. Robotic surgical education: a systematic approach to training urology residents to perform robotic-assisted laparoscopic radical prostatectomy. Urology. 2006;68(1):75-79. doi:10.1016/j.urology.2006.01.057.

42. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative “bench station” examination. The American Journal of Surgery. 1997;173(3):226-230.

43. 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. doi:10.1016/j.urology.2014.09.017.

44. 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. European Urology. November 2015. doi:10.1016/j.eururo.2015.10.048.

45. L LA PIETRA LCLMRQSB. Medical errors and clinical risk management: state of the art. March 2006:1-8.

46. 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. The Journal of Urology. November 2016. doi:10.1016/j.juro.2016.11.100.

47. Liu M, Purohit S, Mazanetz J, Allen W, Kreaden US, Curet M. Assessment of Robotic Console Skills (ARCS): construct validity of a novel global rating scale for technical skills in robotically assisted surgery. Surgical Endoscopy. 2017;94(5):373. doi:10.1007/s00464-017-5694-7.

48. McVey R, Goldenberg MG, Bernardini M, et al. Baseline Laparoscopic Skill May Predict Baseline Robotic Skill and Early Robotic Surgery Learning Curve. J Endourol. February 2016:end.2015.0774. doi:10.1089/end.2015.0774.

49. Goh AC, Aghazadeh MA, Mercado MA, et al. Multi-Institutional Validation of Fundamental Inanimate Robotic Skills Tasks. The Journal of Urology. 2015;194(6):1751-1756. doi:10.1016/j.juro.2015.04.125.

50. Kim JY, Kim SB, Pyun JH, et al. Concurrent and predictive validation of robotic simulator Tube 3 module. Korean J Urol. 2015;56(11):756-761. doi:10.4111/kju.2015.56.11.756.

51. Hinata N, Iwamoto H, Morizane S, et al. Dry box training with three-dimensional vision for the assistant surgeon in robot-assisted urological surgery. Int J Urol. 2013;20(10):1037-1041. doi:10.1111/iju.12101.

52. Arain NA, Dulan G, Hogg DC, et al. Comprehensive proficiency-based inanimate training for robotic surgery: reliability, feasibility, and educational benefit. Surg Endosc. 2012;26(10):2740-2745. doi:10.1007/s00464-012-2264-x.

53. Lendvay TS, Brand TC, White L, et al. Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: a prospective randomized controlled study. Journal of the American College of Surgeons. 2013;216(6):1181-1192. doi:10.1016/j.jamcollsurg.2013.02.012.

54. Lee JY, Mucksavage P, Kerbl DC, Huynh VB, Etafy M, McDougall EM. Validation study of a virtual reality robotic simulator--role as an assessment tool? The Journal of Urology. 2012;187(3):998-1002. doi:10.1016/j.juro.2011.10.160.

55. Tausch TJ, Kowalewski TM, White LW, McDonough PS, Brand TC, Lendvay TS. Content and construct validation of a robotic surgery curriculum using an electromagnetic instrument tracker. The Journal of Urology. 2012;188(3):919-923. doi:10.1016/j.juro.2012.05.005.

56. Dulan G, Rege RV, Hogg DC, et al. Proficiency-based training for robotic surgery: construct validity, workload, and expert levels for nine inanimate exercises. Surg Endosc. 2012;26(6):1516-1521. doi:10.1007/s00464-011-2102-6.

57. Stegemann AP, Ahmed K, Syed JR, et al. Fundamental skills of robotic surgery: a multi-institutional randomized controlled trial for validation of a simulation-based curriculum. Urology. 2013;81(4):767-774. doi:10.1016/j.urology.2012.12.033.

58. Davis JW, Kamat A, Munsell M, Pettaway C, Pisters L, Matin S. Initial experience of teaching robot-assisted radical prostatectomy to surgeons-in-training: can training be evaluated and standardized? BJU Int. 2010;105(8):1148-1154. doi:10.1111/j.1464-410X.2009.08997.x.

59. Amirian MJ, Lindner SM, Trabulsi EJ, Lallas CD. Surgical suturing training with virtual reality simulation versus dry lab practice: an evaluation of performance improvement, content, and face validity. J Robotic Surg. 2014;8(4):329-335. doi:10.1007/s11701-014-0475-y.

60. Foell K, Finelli A, Yasufuku K, et al. Robotic surgery basic skills training: Evaluation of a pilot multidisciplinary simulation-based curriculum. Can Urol Assoc J. 2013;7(11-12):430-434. doi:10.5489/cuaj.222.

61. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. Annals of Surgery. 2004;240(3):518–25–discussion525–8.

62. Menhadji A, Abdelshehid C, Osann K, et al. Tracking and assessment of technical skills acquisition among urology residents for open, laparoscopic, and

robotic skills over 4 years: is there a trend? J Endourol. 2013;27(6):783-789. doi:10.1089/end.2012.0633.

63. Moglia A, Ferrari V, Morelli L, Ferrari M, Mosca F, Cuschieri A. A Systematic Review of Virtual Reality Simulators for Robot-assisted Surgery. European Urology. September 2015. doi:10.1016/j.eururo.2015.09.021.

64. Wiener S, Haddock P, Shichman S, Dorin R. Construction of a Urologic Robotic Surgery Training Curriculum: How Many Simulator Sessions Are Required for Residents to Achieve Proficiency? J Endourol. 2015;29(11):1289-1293. doi:10.1089/end.2015.0392.

65. 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. 2017;7(5):431–11. doi:10.1007/s00464-017-5634-6.

66. Noureldin YA, Elkoushy MA, Aloosh M, Carrier S, Elhilali MM, Andonian S. Objective Structured Assessment of Technical Skills for the Photoselective Vaporization of the Prostate Procedure: A Pilot Study. J Endourol. 2016;30(8):923-929. doi:10.1089/end.2016.0270.

67. Meier M, Horton K, John H. Da Vinci© Skills Simulator™: is an early selection of talented console surgeons possible? J Robotic Surg. 2016;10(4):289-296. doi:10.1007/s11701-016-0616-6.

68. Kelly DC, Margules AC, Kundavaram CR, et al. Face, content, and construct validation of the da Vinci Skills Simulator. Urology. 2012;79(5):1068-1072. doi:10.1016/j.urology.2012.01.028.

69. Finnegan KT, Meraney AM, Staff I, Shichman SJ. da Vinci Skills Simulator Construct Validation Study: Correlation of Prior Robotic Experience With

Overall Score and Time Score Simulator Performance. Urology. 2012;80(2):330-336. doi:10.1016/j.urology.2012.02.059.

70. Face, Content and Construct Validity of a Novel Robotic Surgery Simulator. JURO. 2011;186(3):1019-1025. doi:10.1016/j.juro.2011.04.064.

71. Foell K, Furse A, Honey RJD, Pace KT, Lee JY. Multidisciplinary validation study of the da Vinci Skills Simulator: educational tool and assessment device. J Robotic Surg. 2013;7(4):365-369. doi:10.1007/s11701-013-0403-6.

72. Brown K, Mosley N, Tierney J. Battle of the bots: a comparison of the standard da Vinci and the da Vinci Surgical Skills Simulator in surgical skills acquisition. Journal of Robotic Surgery. 2017;11(2):159-162. doi:10.1007/s11701-016-0636-2.

73. Yang K, Zhen H, Hubert N, Perez M, Wang XH, Hubert J. From dV-Trainer to Real Robotic Console: The Limitations of Robotic Skill Training. J Surg Educ. April 2017. doi:10.1016/j.jsurg.2017.03.006.

74. Lyons C, Goldfarb D, Jones SL, et al. Which skills really matter? proving face, content, and construct validity for a commercial robotic simulator. Surg Endosc. 2013;27(6):2020-2030. doi:10.1007/s00464-012-2704-7.

75. Mark JR, Kelly DC, Trabulsi EJ, Shenot PJ, Lallas CD. The effects of fatigue on robotic surgical skill training in Urology residents. J Robotic Surg. 2014;8(3):269-275. doi:10.1007/s11701-014-0466-z.

76. Yamany T, Woldu SL, Korets R, Badani KK. Effect of postcall fatigue on surgical skills measured by a robotic simulator. J Endourol. 2015;29(4):479-484. doi:10.1089/end.2014.0349.

77. Liss MA, Kane CJ, Chen T, Baumgartner J, Derweesh IH. Virtual reality suturing task as an objective test for robotic experience assessment. BMC Urology. 2015;15(1):63-67. doi:10.1186/s12894-015-0051-4.

78. Alzahrani T, Haddad R, Alkhayal A, 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-E529. doi:10.5489/cuaj.419.

79. Hassan SO, Dudhia J, Syed LH, et al. Conventional Laparoscopic vs Robotic Training: Which is Better for Naive Users? A Randomized Prospective Crossover Study. J Surg Educ. 2015;72(4):592-599. doi:10.1016/j.jsurg.2014.12.008.

80. Liss MA, Abdelshehid C, Quach S, et al. Validation, correlation, and comparison of the da Vinci trainer(™) and the daVinci surgical skills simulator(™) using the Mimic(™) software for urologic robotic surgical education. J Endourol. 2012;26(12):1629-1634. doi:10.1089/end.2012.0328.

81. Song PH, Ko YH. The Surgical Skill of a Novice Trainee Manifests in Time-Consuming Exercises of a Virtual Simulator Rather Than a Quick-Finishing Counterpart: A Concurrent Validity Study Using an Urethrovesical Anastomosis Model. J Surg Educ. 2016;73(1):166-172. doi:10.1016/j.jsurg.2015.08.010.

82. Tanaka A, Graddy C, Simpson K, Perez M, Truong M, Smith R. Robotic surgery simulation validity and usability comparative analysis. Surg Endosc. November 2015:1-10. doi:10.1007/s00464-015-4667-y.

83. Lerner MA, Ayalew M, Peine WJ, Sundaram CP. Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system? J Endourol. 2010;24(3):467-472. doi:10.1089/end.2009.0190.

84. Lee JY, Mucksavage P, Canales C, McDougall EM, Lin S. High Fidelity Simulation Based Team Training in Urology: A Preliminary Interdisciplinary Study of Technical and Nontechnical Skills in Laparoscopic Complications Management. JURO. 2012;187(4):1385-1391. doi:10.1016/j.juro.2011.11.106.

85. 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(5):804-811. doi:10.1111/bju.13710.

86. Kang SG, Cho S, Kang SH, et al. The Tube 3 module designed for practicing vesicourethral anastomosis in a virtual reality robotic simulator: determination of face, content, and construct validity. Urology. 2014;84(2):345-350. doi:10.1016/j.urology.2014.05.005.

87. Perrenot C, Perez M, Tran N, et al. The virtual reality simulator dV-Trainer(®) is a valid assessment tool for robotic surgical skills. Surg Endosc. 2012;26(9):2587-2593. doi:10.1007/s00464-012-2237-0.

88. Kenney PA, Wszolek MF, Gould JJ, Libertino JA, Moinzadeh A. Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. Urology. 2009;73(6):1288-1292. doi:10.1016/j.urology.2008.12.044.

89. Lendvay TS, Casale P, Sweet R, Peters C. Initial validation of a virtual-reality robotic simulator. J Robotic Surg. 2008;2(3):145-149. doi:10.1007/s11701-008-0099-1.

90. Sethi AS, Peine WJ, Mohammadi Y, Sundaram CP. Validation of a novel virtual reality robotic simulator. J Endourol. 2009;23(3):503-508. doi:10.1089/end.2008.0250.

91. Schommer E, Patel VR, Mouraviev V, Thomas C, Thiel DD. Diffusion of Robotic Technology Into Urologic Practice has Led to Improved Resident Physician Robotic Skills. J Surg Educ. 2017;74(1):55-60. doi:10.1016/j.jsurg.2016.06.006.

92. Ruparel RK, Taylor AS, Patel J, et al. Assessment of virtual reality robotic simulation performance by urology resident trainees. J Surg Educ. 2014;71(3):302-308. doi:10.1016/j.jsurg.2013.09.009.

93. FACS TSLM, FACS TCBM, PhC LWBH, et al. Virtual Reality Robotic Surgery Warm-Up Improves Task Performance in a Dry Laboratory Environment: A&nbsp;Prospective Randomized Controlled Study. Journal of the American College of Surgeons. 2013;216(6):1181-1192. doi:10.1016/j.jamcollsurg.2013.02.012.

94. Lendvay TS, Casale P, Sweet R, Peters C. VR robotic surgery: randomized blinded study of the dV-Trainer robotic simulator. Stud Health Technol Inform. 2008;132:242-244.

95. Whittaker G, Aydin A, Raison N, et al. Validation of the RobotiX Mentor Robotic Surgery Simulator. Journal of Endourology. 2016;30(3):338-346. doi:10.1089/end.2015.0620.

96. Chowriappa AJ, Shi Y, Raza SJ, et al. Development and validation of a composite scoring system for robot-assisted surgical training--the Robotic Skills Assessment Score. J Surg Res. 2013;185(2):561-569. doi:10.1016/j.jss.2013.06.054.

97. Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G, Chandrasekhar R, Wilding GE, Guru KA. Face validation of a novel robotic surgical simulator. Urology. 2010;76(2):357-360. doi:10.1016/j.urology.2009.11.069.

98. Kesavadas T, Kumar A, Srimathveeravalli G. Efficacy of Robotic Surgery SImulator (RoSS) for the Davinci® Surgical System. The Journal of …; 2009.

99. Guru KA, Baheti A, Kesavadas T, Kumar A. In-Vivo Videos Enhance Cognitive Skills for Da Vinci® Surgical System. The Journal of …; 2009.

100. Gavazzi A, Bahsoun AN, Van Haute W, et al. Face, content and construct validity of a virtual reality simulator for robotic surgery (SEP Robot). annals. 2011;93(2):152-156. doi:10.1308/003588411X12851639108358.

101. 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(3):518-523. doi:10.1111/j.1464-410X.2012.11204.x.

102. Balasundaram I, Aggarwal R, Darzi A. Short-phase training on a virtual reality simulator improves technical performance in tele-robotic surgery. Int J Med Robot. 2008;4(2):139-145. doi:10.1002/rcs.181.

103. McDonough PS, Tausch TJ, Peterson AC, Brand TC. Initial validation of the ProMIS surgical simulator as an objective measure of robotic task performance. J Robotic Surg. 2011;5(3):195-199. doi:10.1007/s11701-011-0256-9.

104. Jonsson MN, Mahmood M, Askerud T, et al. ProMIS ™Can Serve as a da Vinci ®Simulator—A Construct Validity Study. Journal of Endourology. 2011;25(2):345-350. doi:10.1089/end.2010.0220.

105. Chandra V, Nehra D, Parent R, et al. A comparison of laparoscopic and robotic assisted suturing performance by experts and novices. Surgery. 2010;147(6):830-839. doi:10.1016/j.surg.2009.11.002.

106. 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. Journal of Endourology. 2015;29(11):1295-1301. doi:10.1089/end.2015.0191.

107. Kumar R, Jog A, Malpani A, et al. Assessing system operation skills in robotic surgery trainees. Int J Med Robot. 2012;8(1):118-124. doi:10.1002/rcs.449.

108. Yang K, Perez M, Hossu G, Hubert N, Perrenot C, Hubert J. “Alarm-corrected” ergonomic armrest use could improve learning curves of novices on robotic simulator. Surg Endosc. 2017;31(1):100-106. doi:10.1007/s00464-016-4934-6.

109. Lee JY, Mucksavage P, Sundaram CP, McDougall EM. Best Practices for Robotic Surgery Training and Credentialing. The Journal of Urology. 2011;185(4):1191-1197. doi:10.1016/j.juro.2010.11.067.

110. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Health Sci Educ Theory Pract. 2014;19(2):233-250. doi:10.1007/s10459-013-9458-4.

111. MD JRPI. Assessment of Competence. Surgical Clinics of NA. 2016;96(1):15-24. doi:10.1016/j.suc.2015.08.008.

112. Canada RCOPASO. Competence by Design: Reshaping Canadian Medical Education. March 2014:1-141.

113. Hammond L, Ketchum J, Schwartz BF. Accreditation Council on Graduate Medical Education Technical Skills Competency Compliance: Urologic Surgical Skills. Journal of the American College of Surgeons. 2005;201(3):454-457. doi:10.1016/j.jamcollsurg.2005.05.002.

Table-1.1 Validity evidence for assessments of technical skill

Assessment Method	**Instrument Description**	**Domains Assessed**	**Number of Studies, Primary Assessment Method**	**Number of Participants, Primary Assessment Method**	**Number of Studies, Secondary Assessment Method**	**Content***	**Response Process***	**Internal Structure***	**Relationship to Other Variables***	**Consequences of Testing***
GEARS	Robotic-specific GRS; expansion of GOALS with expert consensus	Depth perception Bimanual dexterity Efficiency Force Sensitivity Autonomy Robotic control	18	569	2	17	11	11 IRR: 0.38-0.92 (M=0.80)	12	1. Scores used to determine ranking 2. GEARS score predicts surgical outcome
OSATS	GRS; developed initially for assessing basic surgical skills in OSCE examination	Respect for Tissue Time and Motion Instrument Handling Knowledge of Instruments Flow of Procedure Use of Assistants Knowledge of Procedure	7	345	1	9	4	3 IRR: 0.84-0.91 (M=0.87)	8	0

Assessment Method	Instrument Description	Domains Assessed	Number of Studies, Primary Assessment Method	Number of Participants, Primary Assessment Method	Number of Studies, Secondary Assessment Method	Content*	Response Process*	Internal Structure*	Relationship to Other Variables*	Consequences of Testing*
GOALS	GRS; developed to assess laparoscopic skill	Depth perception Bimanual dexterity Efficiency Tissue Handling Autonomy	3	72	1	4	3	2 IRR: 0.66-0.80	3	0
R-OSATS	GRS; four dry-lab exercise-specific scale that combines elements of GOALS and OSATS	Depth Perception Force Sensitivity Dexterity Efficiency	1	105	0	1	1	1 IRR: 0.79	1	0
PACE	Procedure-Specific GRS for RARP	Anchored Likert Scale Across 7 Operative Steps	1	56	0	1	1	1 IRR: 0.4-0.8	1	0

Assessment Method	Instrument Description	Domains Assessed	Number of Studies, Primary Assessment Method	Number of Participants, Primary Assessment Method	Number of Studies, Secondary Assessment Method	Content*	Response Process*	Internal Structure*	Relationship to Other Variables*	Consequences of Testing*
ARCS	Robotic-specific GRS developed by Intuitive Surgical technician-trainers	Dexterity Optimizing Field of View Instrument Visualization Optimizing Workspace Force Sensitivity and Control Basic Energy Pedal Skills	1	15	0	1	1	1 IRR: 0.52-0.81	1	0
RACE	Task-Specific GRS developed to evaluate urethrovesical anastomosis performance	Needle Positioning Needle Entry Needle Driving & Tissue Trauma Suture Placement Tissue Approximation Knot Tying	2	40	0	2	1	1 IRR: 0.55-0.62	1	0

Assessment Method	**Instrument Description**	**Domains Assessed**	**Number of Studies, Primary Assessment Method**	**Number of Participants, Primary Assessment Method**	**Number of Studies, Secondary Assessment Method**	**Content**[*]	**Response Process**[*]	**Internal Structure**[*]	**Relationship to Other Variables**[*]	**Consequences of Testing**[*]
RARP Assessment Score	Prostatectomy-Specific Assessment based on HFMEA analysis	Operative steps broken down into sub-steps with hazard categories assigned for modular introduction to RARP	1	15	0	1	1	1 Kappa range - 0.241 – 0.2 Significant agreement on 2/27 steps	1	0

[*]based on Messick's Framework of Validity

Table 1.2 Validity evidence for computer-based virtual reality assessments

Assessment Method	**Instrument Description**	**Domains Assessed**	**Number of Studies, Primary Assessment Method**	**Number of Participants, Primary Assessment Method**	**Number of Studies, Secondary Assessment Method**	**Content**[*]	**Response Process**[*]	**Internal Structure**[*]	**Relationship to Other Variables**[*]	**Consequences of Testing**[*]

Assessment Method	**Instrument Description**	**Domains Assessed**	**Number of Studies, Primary Assessment Method**	**Number of Participants, Primary Assessment Method**	**Number of Studies, Secondary Assessment Method**	**Content**[*]	**Response Process**[*]	**Internal Structure**[*]	**Relationship to Other Variables**[*]	**Consequences of Testing**[*]
dV-Trainer/ MdVT	Computer-Generated metrics developed by Mimic Simulation	Time Economy of Motion Drops Instrument Collisions Excessive Instrument Force Instruments Out of View Master Workspace Range	12	525	1	12	2	0	8	0
dVSS	Computer-Generated metrics developed by Intuitive Surgical	Camera targeting Energy switching Threading rings Dots and Needles Ring and rail	23	697	3	26	12	0	21	1. dVSS scores predict GEARS score in OR 2. dVSS scores predict performance on dry-lab tasks using

Assessment Method	Instrument Description	Domains Assessed	Number of Studies, Primary Assessment Method	Number of Participants, Primary Assessment Method	Number of Studies, Secondary Assessment Method	Content*	Response Process*	Internal Structure*	Relationship to Other Variables*	Consequences of Testing*
										robotic console
RoSS (RSA-Score)	Computer-Generated metrics developed by the University of Buffalo and the Roswell Cancer Institute	Task Time Safety in Operative Field Economy Bimanual Dexterity Critical Errors	2	57	0	2	0	1 Internal Domain Consistency 0.01-0.98	1	0

Assessment Method	Instrument Description	Domains Assessed	Number of Studies, Primary Assessment Method	Number of Participants, Primary Assessment Method	Number of Studies, Secondary Assessment Method	Content*	Response Process*	Internal Structure*	Relationship to Other Variables*	Consequences of Testing*
SEP	Simulator developed in the Netherlands		2	63	0	2	0	1 IRR: 0.73	1	0
RobotiX	Computer-Generated metrics developed by Simbionix Products	Fundamentals of Robotic Surgery and Robotic Suturing Modules	1	46	0	1	0	0	1	0
ProMIS	Adapted from Laparoscopic Training System from Haptica (Ireland)	Peg Transfer Precision Cut Intracorporeal Knot	3	73	0	3	0	0	3	0

*based on Messick's Framework of Validity

Table-2 Novel methods of assessing robotic skill

Assessment Method	Description of Innovation	Levels of Training	Setting of Assessment	Advantages of Method
Crowdsourced Assessments	Enlists large numbers of people via an internet platform to complete assessments of technical skill	Medical Students Residents Fellows Staff	Dry-Lab Simulation Wet-Lab Operating Room	Rapid, high volume assessments of video High interrater reliability statistics
Machine Learning	Automated analysis of master workspace adjustment, camera manipulation, unsafe motion and collisions	Residents Fellows	Dry-Lab	Automated analysis of surgeon psychometrics Excellent classification accuracy Potential for real-time, high reliability assessment of performance
Contact Vibrations	Use of contact vibrations, applied force, and time to completion as measures of clinical skill	Staff	Dry-Lab	Improvement classification accuracy of a global rating scale assessment of technical skill
Armrest Load	Use of a pressure surveillance system to detect armrest load on the robotic console	Medical Students	Simulation	Use of pressure-alarm in training can improve ergonomic positioning in novice surgeons Potential for shortening of learning curve in novice trainees

Table-3 Description of high quality evidence (MERSQI ≥ 14). Arranged in ascending order of MERSQI score

Study	**Trainees**	**Setting, Type of Assessment**	**Assessment summary**	**Measurement Tool**	**Conclusion**	**MERSQI**
Vlaovic et al. (2008)	101 T	Dry, TS	5-day laparoscopic training program. Includes 2-3 hrs of lectures, daily practice on pelvic trainers and VR simulators, and training on porcine models and human cadavers. Assessed ring transfer, suture threading, cutting, and suturing by expert examiner	OSATS	Post-course robotic performance was significantly improved ($p < 0.001$)	14
Davis et al. (2010)	3 R 4 F	OR, TS	Standardized method of evaluating performance in robot-assisted radical prostatectomy using time, autonomy scale and end-product assessment by expert surgeons	Time, quality of results relative to staff, short term patient outcomes	Time to completion was longer for trainee’s vs staff ($p < 0.001$), basic vs advanced tissue dissection and suturing. No increase in adverse short-term outcomes was observed	14

Study	Trainees	Setting, Type of Assessment	Assessment summary	Measurement Tool	Conclusion	MERSQI
Kumar et al. (2012)	6 novice 2 expert	Dry, TS	Support Vector Machines (SVM) to classify expert-novice operational skills. Assessed master workspace adjustment, camera manipulation skills, unsafe motion and collisions by computer for manipulation, suturing, transection, and dissection	Support Vector Machines (SVM)	Model correctly classified 91.7% for master workspace and 88.2% for camera manipulation	14
Foell et al. (2013)	29 R 16 F 8 S	VR/Dry, TS	Participants included urology, obstetrics and gynecology, and thoracic surgery. Assessed Camera Targeting 1, Peg Board 1, Match Board 1, Thread the Rings, Suture Sponge 1, Ring Walk 2, and Peg Board 2 by dVSS, and compared to dry-lab performance on robotic console	dVSS metrics, time/number of errors	Performance on dVSS modules had moderate-strong correlation with time/error assessment on robotic console in dry-lab setting	14

Study	Trainees	Setting, Type of Assessment	Assessment summary	Measurement Tool	Conclusion	MERSQI
Yamany et al. (2015)	13 R	Dry, TS	Effect of 24-hr call on suturing performance of residents with or without prior robotic simulator experience. Participants included urology and general surgery. Assessed time to completion of exercise, needle loading, knot tying by dVSS	dVSS metrics	Time to completion, needle loading, and knot tying were significantly increased postcall ($p < 0.05$). Prior simulator experience did not have significant benefits in postcall performance ($p < 0.05$)	14
Whitehurst et al. (2015)	7 R 8 F 5 S	dV-Trainer/Dry/Wet (swine), TS	Compared robotic performance between training in a VR or dry lab setting. Participants included gynecology, urogynecology, gynecologic oncology, reproductive endocrinology, and urology. Assessed cystotomy closure on swine model by blinded expert surgeons	dV-Trainer metrics, GEARS	Training modalities did not differ significantly: 2.83 ± 0.66 for VR cohort, 2.96 ± 0.77 for dry cohort, $p = 0.690$	14

Study	Trainees	Setting, Type of Assessment	Assessment summary	Measurement Tool	Conclusion	MERSQI
McVey et al. (2016)	11 R 21 F	Box-Trainer, TS	Effect of baseline laparoscopic skill on robotic skill before and after robotic surgery basic skills training course. Participants included urology, gynecology, thoracic surgery, and general surgery. Assessed by two blinded content experts using Likert scale global rating score	Time, number of errors	Baseline laparoscopic intracorporeal suturing and knot tying (ISKT) performance strongly correlated with robotic performance ($p = 0.01$ for peg transfer, $p < 0.01$ for ISKT). IRR = 0.9	14
Chowriappa et al. (2013)	15 novice 12 expert	VR, TS	Assessed fourth arm control, coordinated tool control, ball placement, and needle handling and exchange by RoSS simulator	RSA-Score	Expert cohort performed significantly across all tasks: $p = 0.002$ for fourth arm control, $p < 0.001$ for coordinated tool control, $p < 0.001$ for ball placement, $p < 0.001$ for needle handling and exchange	14.5

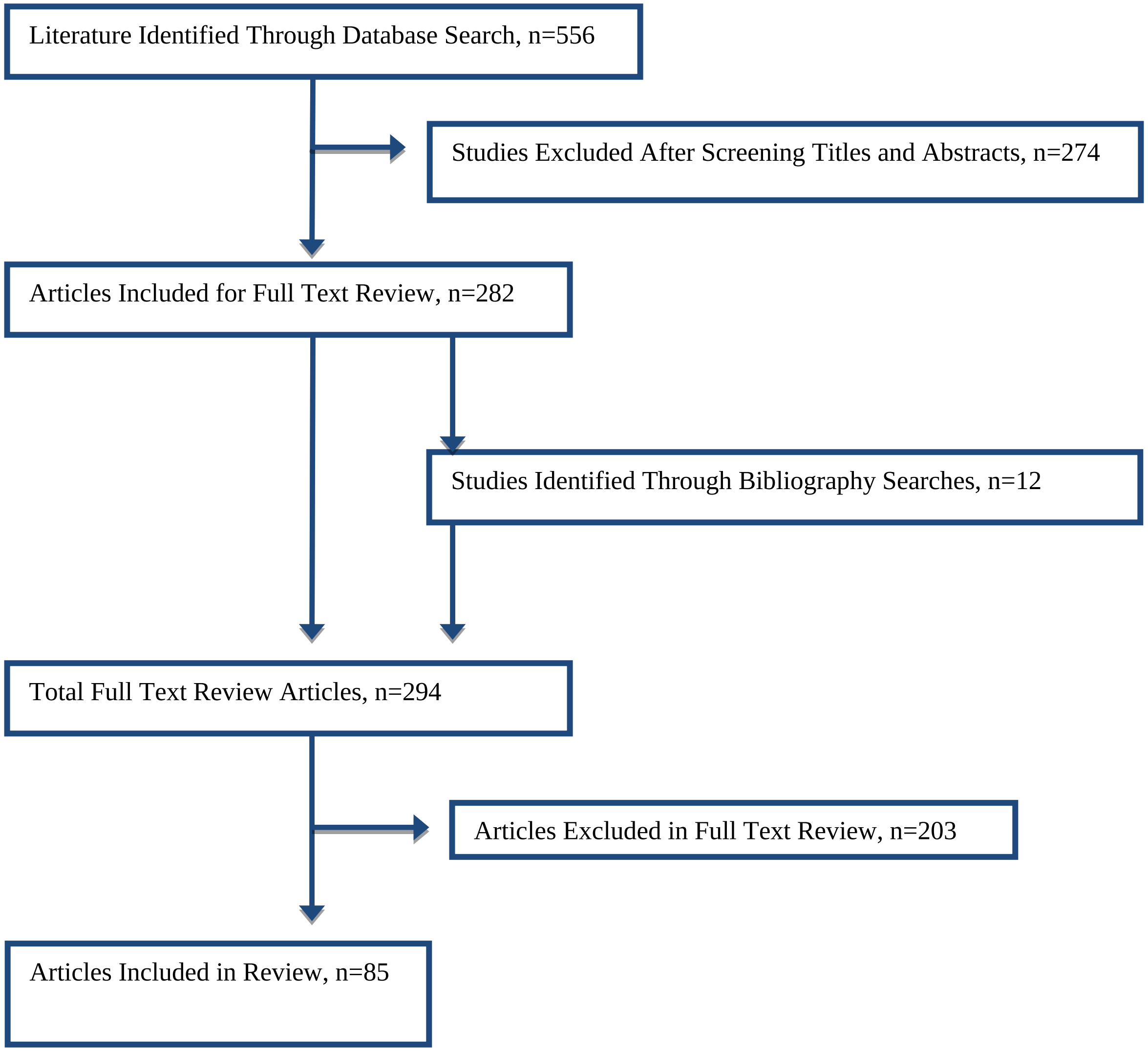
Study	Trainees	Setting, Type of Assessment	Assessment summary	Measurement Tool	Conclusion	MERSQI
Hung et al. (2015)	15 novice 13 intermediate 14 expert	AR/VR, TS	Developed simulation platform for robotic partial nephrectomy. Includes augmented reality content and virtual reality renorrhaphy. Assessed by blinded expert reviewer	dV-Trainer metrics, GEARS	Simulation platform demonstrated strong face, content, and construct validity. Virtual reality renorrhaphy performance correlated significantly with porcine model ($r = 0.8$, $p < 0.0001$)	14.5
Schommer et al. (2017)	34 R	dV-Trainer, TS	Compared access to robotic technology to robotic skill between residents attending a training course in 2012 and 2015. Assessed Camera Targeting 2, Energy Dissection 1, Needle Targeting, and Peg Board 1 by dV-Trainer	dV-Trainer metrics	Robotic performance was significantly better in the 2015 cohort than 2012 ($p < 0.001$). Access to robot console correlated with better scores in Camera Targeting 2 ($p = 0.02$) and Peg Board ($p = 0.04$)	14.5

Study	Trainees	Setting, Type of Assessment	Assessment summary	Measurement Tool	Conclusion	MERSQI
Raison et al. (2017)	102 R 121 S	dV-Trainer, TS	Set benchmark scores to achieve competency in robot skills. Assessed basic (Pick and Place, Camera Targeting 1, Peg Board 1) and advanced (Thread the Rings 1, Suture Sponge) tasks by dV-Trainer	dV-Trainer metrics	Using a benchmark score of 75% of the mean expert score, novice trainees achieved competency in basic but not advanced tasks. Intermediate trainees achieved competency in basic tasks and Suture Sponge	14.5
Xu et al. (2016)	11 Robotic-experienced 7 Laparoscopic-Experienced 9 Control	Xperience Team-Trainer (XTT)	Establish initial validity evidence for a team-based robotic surgery simulator, including bedside assistant involvement. Evaluated simulation performance as assistant and console surgeon using the XTT	XTT Metrics, Modified GOALS	Demonstrated that scores on XTT correlate with both robotic experience and performance on the console. The robotic and laparoscopic experienced surgeons outperformed controls in all exercises.	14.5
Stegemann et al. (2013)	9 MS 26 R 10 F 8 S	Box trainer, TS	Provide validity evidence, demonstrating that Fundamental Skills of Robotic Surgery (FSRS) curriculum completion improves performance on tasks completed with actual daVinci console in simulation setting	Number of errors, camera/clutch use	Although no differences between study arms, control group showed significant improvement from baseline on repeat daVinci console scores when allowed to crossover into FSRS arm	14.5

Study	Trainees	Setting, Type of Assessment	Assessment summary	Measurement Tool	Conclusion	MERSQI
Lendvay et al. (2013)	27 R 24 S	VR/Dry, TS	Effect of VR warm-up on robotic performance in similar and dissimilar tasks. Participants included general surgery, urology, and gynecology. Assessed rotating rocking pegboard and intracorporeal suturing by computer	Time, cognitive and technical errors, tool path length, economy of motion	Warm-up cohort performed significantly better in time ($p = 0.001$) and path length ($p = 0.014$) for similar tasks (rotating rocking pegboard) and significantly better in global technical errors ($p = 0.020$) for dissimilar tasks (intracorporeal suturing)	15
Tarr et al. (2014)	99 R	Dry, TS	Compared robotic performance before and after an unstructured or structured robotic training curriculum. Structured curriculum included specific instructions and goal times to achieve before proceeding to the next task. Participants included gynecology and urology. Assessed manipulation, transection, knot tying, and suturing by expert examiner	OSATS	Structured cohort performed significantly better in transection ($p < 0.05$), while unstructured cohort performed significantly better in knot tying ($p < 0.05$). No significant differences were observed in manipulation and suturing	15

MS – medical student, R – resident, F – fellow, S – staff, T – trainee, TS – technical skill, IRR – inter-rater reliability

Figure-1 PRISMA flow chart


Literature Identified Through Database Search, n=556
Studies Excluded After Screening Titles and Abstracts, n=274
Articles Included for Full Text Review, n=282
Studies Identified Through Bibliography Searches, n=12
Total Full Text Review Articles, n=294
Articles Excluded in Full Text Review, n=203
Articles Included in Review, n=85

Author/s:
Goldenberg, MG;Lee, JY;Kwong, JCC;Grantcharov, TP;Costello, A

Title:
Implementing assessments of robot-assisted technical skill in urological education: a systematic review and synthesis of the validity evidence

Date:
2018-09

Citation:
Goldenberg, M. G., Lee, J. Y., Kwong, J. C. C., Grantcharov, T. P. & Costello, A. (2018). Implementing assessments of robot-assisted technical skill in urological education: a systematic review and synthesis of the validity evidence. BJU INTERNATIONAL, 122 (3), pp.501-519. https://doi.org/10.1111/bju.14219.

Persistent Link:
http://hdl.handle.net/11343/283769