Title: Measures of Performance and Proficiency in Robotic-Assisted Surgery: A Systematic Review

Authors

C El-Sayed

A Yiu

J Burke

P Vaughan-Shaw

J Todd

P Lin

Z Kasmani

C Munsch

L Rooshenas

M Campbell

S P Bach

Corresponding Author: Charlotte El-Sayed
RCS England/HEE Robotics Research Fellow
University of Birmingham
Email: c.el-sayed@nhs.net
ORCID ID: 000-0002-5994-5645

Word Count: 3061 (excluding tables and figures)

Acknowledgements: The corresponding author C El-Sayed received a research fellowship by the Royal College of Surgeons England and Health Education England.

Statement and Declarations

Funding: The first author received a research grant from RCS England and Health Education England in November 2021 until present to complete the study.

Competing interests: Simon Bach receives a research funding from Intuitive Surgical Inc. USA.

Ethics approval: none required

Author Declarations: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Charlotte El-Sayed, Andrew Yiu, Josh Burke, James Todd, Paul Lin, Zain Kasmani and Peter Vaughan Shaw. The first draft of the manuscript was written by Charlotte El-sayed and Andrew Yiu. Josh Burke, Chris Munsch, James Todd, Paul Lin, Zain Kasmani, Peter Vaughan-Shaw, Leila Rooshenas, Marion Campbell and Simon Bach critically revised the manuscript based on the study's findings. All authors read and approved the final manuscript.

Abstract

Background: Robotic Assisted Surgery (RAS) has seen a global rise in adoption. Despite this, there is not a standardised training curricula nor a standardised measure of performance. We performed a systematic review across the surgical-specialties in RAS and evaluated tools used to assess surgeon's technical performance.

Methods: Using the PRISMA 2020 guidelines, Pubmed, Embase and the Cochrane Library were searched systematically for full texts published on or after January 2020 - January 2022. Observational studies and RCTs were included; review articles and systematic reviews were excluded. The papers' quality and bias score were assessed using the Newcastle Ottawa Score for the observational studies and Cochrane Risk Tool for the RCTs.

Results: The initial search yielded 1189 papers of which 72 fit the eligibility criteria. 27 unique performance metrics were identified. Global assessments were the most common tool of assessment (n=13); the most used was GEARs (Global Evaluative Assessment of Robotic Skills). 11 metrics (42%) were objective tools of performance. Automated performance metrics (APMs) were the most widely used objective metrics whilst the remaining (n=15, 58%) were subjective.

Conclusion: The results demonstrate variation in tools used to assess technical performance in RAS. A large proportion of the metrics are subjective measures which increases the risk of bias amongst users. A standardised objective metric which measures all domains of technical performance from global to cognitive is required. The metric should be applicable to all RAS procedures and easily implementable. Automated performance metrics (APMs) have demonstrated promise in their wide use of accurate measures.

Word Count:247

Key terms: robotic surgery, metric, curriculum, performance, assessment, training

Introduction

We would like to make all surgery as safe as possible for our patients. We know patients who are treated by surgeons judged to have a high level of technical skill have much better outcomes than those treated by surgeons judged to have a lower level of skill [1]. It should be a simple matter to ensure all surgeons are well trained and receive regular feedback on their performance in the operating theatre. Unfortunately, this is currently not the case and we continue to recognise wide variation in the quality of surgical care that patients receive [2]. To date, quality improvement initiatives have addressed key components of the care pathway that surround surgery, while conduct during the operation is relatively unexplored [3]. This is unfortunate as poor-quality surgery is not only ineffective but exposes patients to high risk of complications and even death. A recent prospective study recorded an average of 20 technical errors and 8 adverse events per case [4]. These technical errors during surgery and associated adverse events consume valuable healthcare resources and may compromise patients' quality of life [5].

Robotic assisted surgery (RAS) is at the forefront of surgical innovation and increasingly used to perform complex operations that pose greatest risk to patients. In 2022, approximately 1,875,000 surgical procedures were performed using the da Vinci Surgical Systems (Intuitive Surgical, USA) [6]. Many competing platforms are emerging including Versius (CMR Surgical, UK) and HUGO (Medtronic Minneapolis, USA)[7, 8]. Compared to conventional laparoscopic techniques, RAS improves patient access to minimally invasive surgery (MIS) while surgical teams benefit from improved visualisation, functionality, and ergonomics. RAS has the capability to support development and implementation of proficiency-based training of novice surgeons and peer-coaching of credentialled surgeons to reduce variation in surgical care and improve patient outcomes [9].

RAS training programmes were developed, and often delivered, by robotic industry providers as opposed to surgical colleges and associations [10]. Despite increasing need [11], there are currently no formal mandatory curricula for robotic training in Europe for any surgical specialty [10, 12]. Higher RAS fellowships serve to complement existing specialty training pathways, however, there is no consensus on how to optimally evaluate and benchmark training outcomes to develop credentialling that incorporates RAS [13]. It is now appropriate to consider how implementation of RAS training within the wider surgical curricula would be optimally evaluated.

It is clear from the current evidence, training in RAS is currently in a state of disarray. RAS training needs to be standardised with quality assurance measures, benchmarking and clear procedural assessment. In order to do so, an appropriate validated metric to assess competency in RAS technical performance is required. At present, the lack of consensus on the appropriate metric(s) with which to assess technical performance has led to a wide degree of variation in the application of these metrics by assessors, limiting comparison and transferrable accreditation. Recent work reviewing clinically relevant performance metrics (CRPMs) in RAS highlighted their use in assessing trainee proficiency, but also found there are no studies at present which correlate metrics with patient outcomes[14]. Furthermore, this work did not assess performance metrics deemed not to be clinically relevant by the authors. An assessment of all metrics used to assess technical performance is needed to provide an accurate understanding of how proficiency is currently assessed and to facilitate with future RAS curriculum development.

The aim of this systematic review is to identify and describe the tools currently established for the measurement of technical performance in RAS across all surgical specialties, and in doing so highlight new areas of technical innovation in the assessment of surgical training. This review will evaluate and compare the currently available metrics to identify the "ideal" metric of assessment.

Methods

Information Sources & Search

This systematic review is reported in accordance with Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) 2020 guidelines. PubMed, Embase and The Cochrane Library were searched systematically for full texts published in English from January 2000 to January 2022 (inclusive) as robotic surgery became prevalent after the start of the millennium. The query (("robotic surgery" or "robotic assisted surgery") OR (robotic surgical procedures [MeSH Terms])) AND ("curriculum" OR " training" OR curriculum [MeSH Terms]) AND ((performance) OR (metric*) OR (clinical competence [MeSH Terms]) OR ("clinical competence")) were used to complete the search. To ensure literature saturation, reference chaining of selected studies were reviewed to identify any missed papers. The review is registered on PROSPERO under ID number CRD42023453528.

Eligibility Criteria

This review included studies which evaluated robotic surgical performance measures used to assess the surgeon's surgical technique. The studies included addressed assessment methods in the operating theatre, laboratory, and virtual reality (VR) settings. Eligible studies included case series, cohort and case control studies and randomised controlled studies (RCTs). Studies were excluded if they were not full text i.e., conference papers isolated abstracts and interim trials or studies which have been published prior to completion. We also excluded systematic reviews and review articles.

Study Selection

Titles and abstracts were screened by CE and AY. Those papers included for full paper review were independently reviewed by CE, AY, PL, JT and ZK. Any ambiguities in the selection process were resolved by the third reviewer, JB.

Outcomes and Prioritisations:

1. To identify and describe the range of metrics available across all surgical disciplines for assessment of technical performance in RAS. Metrics were grouped according to a previously described classification system: (i) global skills assessments, (ii) procedural based, (iii) task based and (iv) cognitive measures [15]. They were further subdivided as subjective or objective. Predefined descriptors included subspecialty training programme (if any), training environment (simulation, laboratory, operating room), type of study and participants were included.
2. To summarise the perceived strengths and weaknesses of each metric.
3. To specify characteristics of an "ideal" metric for RAS technical training assessment.

Data Extraction

Data was compiled in a data extraction sheet in Google Docs. (Google, CA, USA). Extracted data included year of publication, type of study, surgical subspeciality the study was performed in, modality of robotic surgery, the metric used, number of participants and whether construct validity of the metric had been assessed. The quality of the observational studies was assessed using the Newcastle Ottawa Score and the Cochrane risk of bias tool was used to assess RCTs.

Results

The initial search yielded 1189 papers. After the limits 'January 2020 to January 2022' were applied, and after duplicates were removed, 876 results remained to be screened. The 876 abstracts for those references were compiled and screened leaving 98 papers for full review. The final analysis included 72 papers. The PRISMA 2020 flowchart below displays the methodology (Figure 1).

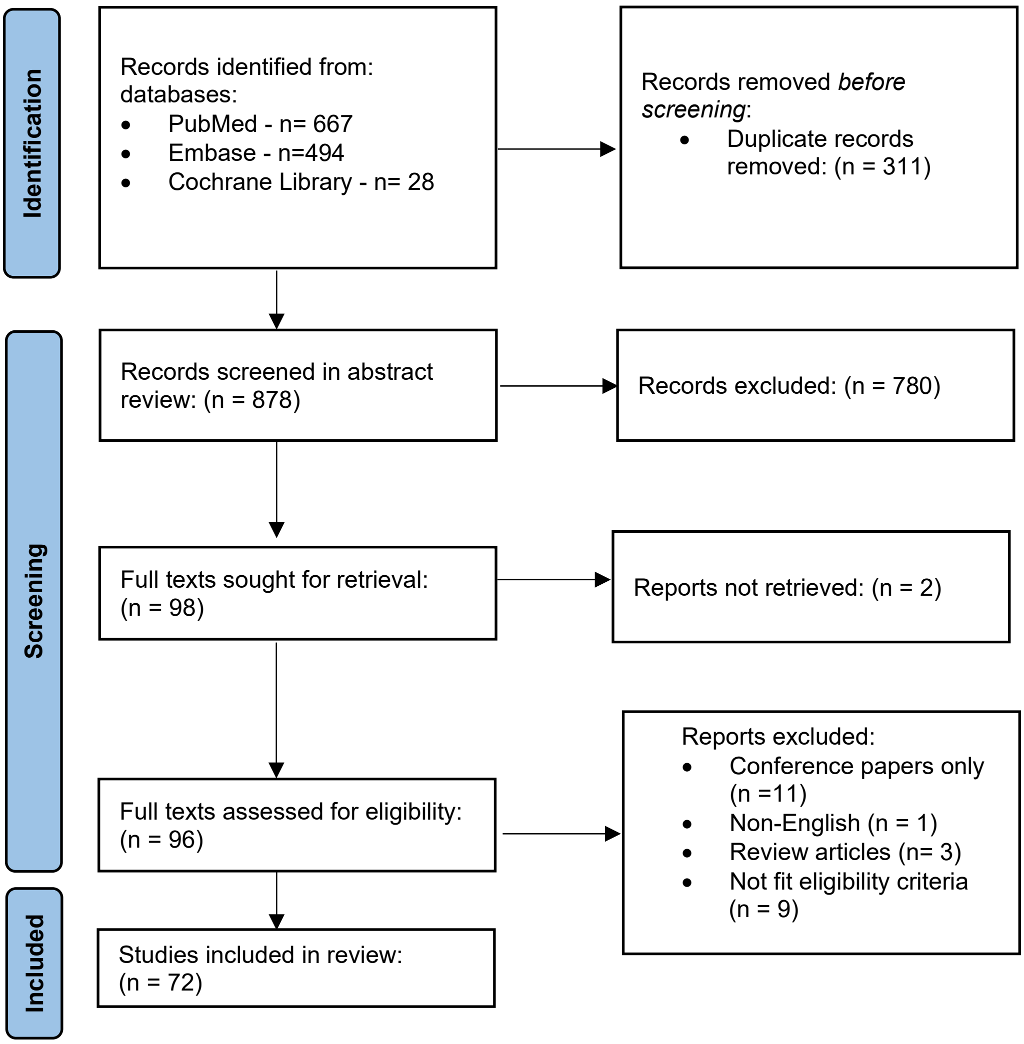

Figure 1: Flow chart of extraction process using PRISMA 2020 statement methodology

<u>Demographics of the studies</u>

On assessing the subspecialities of the papers, the majority of studies were in urology (n=33, 46%) followed by general surgery (n=7, 9%) (Figure 2).

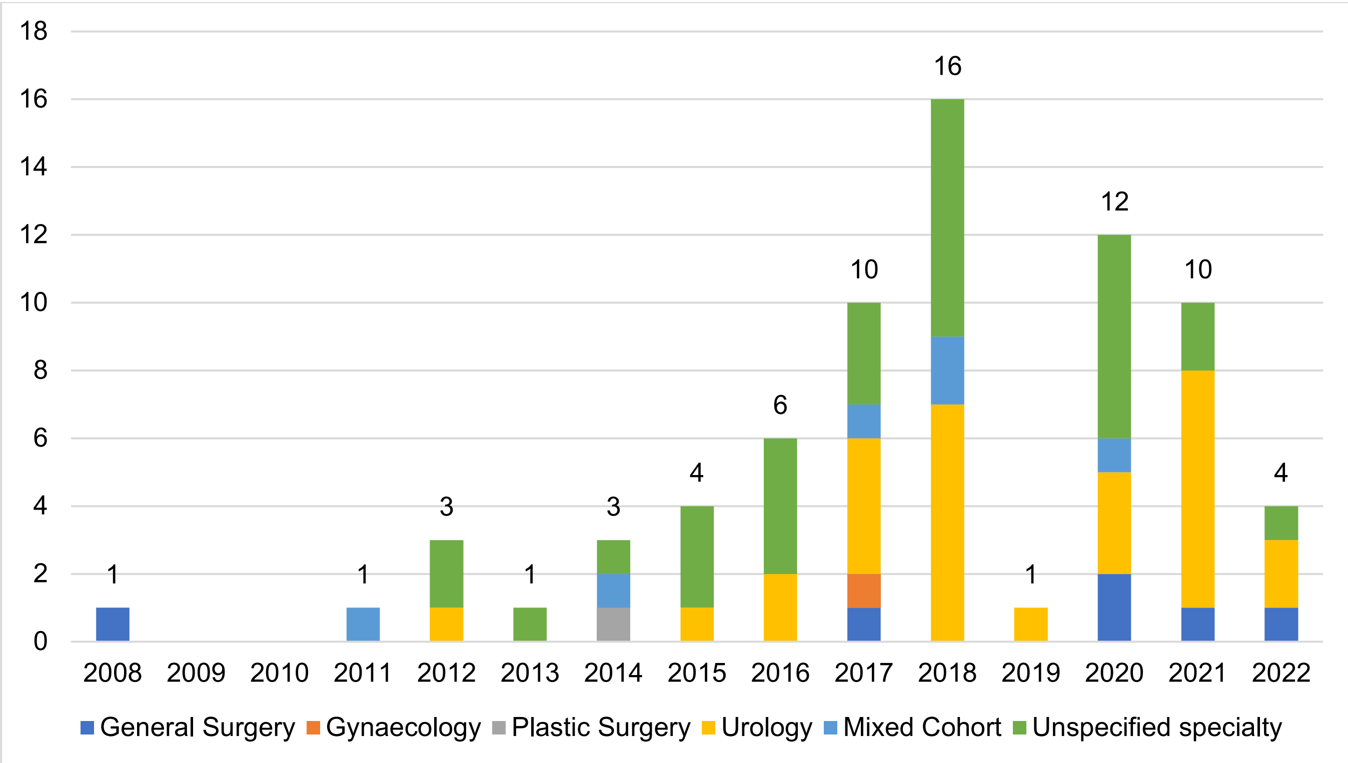

Figure 2: Distribution of studies over time by subspeciality

Twenty-seven unique metrics were identified in total (Table 1). The categorised metrics were subdivided into subjective and objective measures of performance based on the tools used to complete the assessment. Subjective metrics were the most prevalent (n=16, 59%). A description and evaluation of each metric with the referenced papers are included in Table 2.

Table 1: The metrics identified in the review.

	Global Skills Assessment: Tools used to assess a surgeon's general technique focusing on the *basic* robotic surgery performance	**Procedural Based:** Specific metrics which provide surgical skill assessment in a step-by-step approach unique to the procedure.	**Task specific:** metrics specific to a task.	**Cognitive:** metrics which assess the amount of effort being used in the working memory[15].
Subjective Metrics	• Global Evaluative Assessment of Robotic Skills (GEARS) • Objective Structured Assessment Tools (OSATs) • Operative Robotic Index (ORI) • Global Robotic Score (GRS) • Generic Error Rating Tool (GERT) • Robotic-Operative Structured Assessment Tool (R-OSAT) • Structured assessment of robotic microsurgical skills. • Assessment Robotic Console Score (ARCS)	• Robotic Assessment of Radical Prostatectomy (RARP) • Robotic Hysterectomy Assessment Score (RHAS) • Cystectomy Assessment and surgical evaluation (CASE) • Prostatectomy assessment and competency evaluation (PACE) • Crowd sourcing assessment tools (C-SATS)	• Robotic Anastomosis Competency Score (RACE) • assessment of Objective Structured Assessment Tools (aOSATS)	• Nasa Task Load Index (NASA-TLX
Objective Metrics	• Mean Proficiency Index (MPI) • Robotic Skills Assessment Score (RSA) score • Time to completion • Clinically relevant objective metrics of simulators (CROMS)	• Proficiency Based Progression (PBP)	• Camera Metrics • Automated Performance Metrics (APMs)	• Pupillary Measures • Numeric psychomotor test scores - NPMTS • Wisconsin Card Sorting Test (WCST) • Psychomotor Vigilance Task (PVT)

Global Skills assessment

Global skills assessments are the most common domain of metric currently available. Global Evaluative Assessment of Robotic Skills (GEARS) was the most prevalent global skills used in the studies (n=22, 79%)[16-37]. Amongst the objective metrics, Clinically Relevant Objective Measures (CROMS) had the largest sample size, however, the metrics were used only in a simulated environment[38] (Table 2,3).

Procedural Specific Metrics

Of all the metrics identified in this review; the majority were specific to urology; the Prostatectomy Assessment Competency Evaluation (PACE) score, Robotic Assessment Radical Prostatectomy (RARP) score and the Cystectomy Assessment and Surgical Evaluation (CASE) score [39-41]. However, these metrics relied upon subjective measures of assessment. Proficiency Based Progression (PBP) was the only objective measure [42-44] and were the most widely used procedural-specific metric.

Task Specific Metrics

Robotic Anastomosis Competency Evaluation (RACE) was the most prevalent subjective task-based skill used [24, 25, 45-47]. The score is specific to a vesicourethral anastomosis (VUA) only. Objective task-based metrics include both video derived metrics [48] and robotic systems data that describe key elements of intraoperative surgical behaviour including geometric and time-dependent variables of dominant and non-dominant instrument control to yield automated performance metrics (APMs) (42, 45, 48, 51, 54-84). APMs were the most widely used metric in this review. They were used in both the simulated and live operating setting.

Cognitive skills assessment

NASA-TLX was the most common cognitive measure and used in several papers [49-51]. However, NASA-TLX is the only subjective cognitive measure found, as it relies upon self-assessment. Amongst the five objective cognitive measures, pupillary metrics were the most widely used.

Table 2. An evaluation of the metrics identified in the included studies.

Metric	Summary	Number of participants within the review	Advantages	Disadvantages	References
Automated Performance Metrics (APMs)	Computed from robotic instrument motion tracking, events data and surgical videos.	1204	Objective measure of a surgeon's performance. Generalisable to all RAS procedures. Demonstrates construct validity. Widely implementable	Limited evidence regarding clinical correlation between APMs and clinical outcomes. Mainly assessed on simulation models. No assessment of cognitive/decision making.	[26, 29, 32, 35, 46, 48, 49, 52-79]
GEARS	Five-point Likert scale across six domains: depth perception, bimanual dexterity, efficiency, force sensitivity, autonomy, and robotic control.	1072	Widely recognised. Easy to use amongst users. Specific to robotics.	Time consuming to complete. Dependent on Likert scales therefore subject to variation amongst users. No assessment of cognitive/decision making.	[16-37]
Crowd Sourcing Assessment Tool (CSATs)	Use of an autonomous crowd to assess video recordings of a surgical procedure.	476	Reduces subjectivity by using a large crowd of users. Fast assessment tool. Blinded.	Reliant upon GEARs. Requires large numbers of assessors to reduce bias.	[36]
ROSATs	Modified OSATs score specific to robotic surgery. Four domains assessed: (1) depth perception/accuracy (2) force/tissue handling (3) dexterity (4) efficiency (5) robotic arm collisions. Each category is scored using a 5 point Likert scale	151	Easy to use. Applicable to all robotic procedures. Demonstrates construct validity. Can be widely implemented	Dependent on Likert scales therefore subject to variation amongst users.	[80]
NASA-TLX	Six-item questionnaire measuring mental workload. Domains covered include (1)	100	Able to assess subjective workload thereby	Difficult for the respondent to self-	[49-51]

	mental demand (2) physical demand (3) temporal demand (4) performance (5) effort and (6) frustration.		characterising the nature of the mental process in task performance. Widely implemented.	assess cognitive capacity. Limited evidence of its use in robotic training and clinical outcomes. Subjective workload may be influenced by the user's biased judgement.	
Robotic Anastomosis Competency Score (RACE)	Specifically designed for vesicourethral (VUA). The procedure is deconstructed into six key domains. Each domain is given a 3-anchor description matched to a 5-point Likert scale.	110	Demonstrates construct validity – however not statistically significant. Breaks down VUA into key procedural steps.	Limited only to a VUA therefore not applicable to other anastomosis techniques. Only tested on an inanimate model. Dependent on Likert scales therefore subject to variation amongst users.	[25, 45, 47, 81, 82]
MScore Proficiency Index (MPI©)	Single reference number which averages all scores from exercises on the Mimic robotic VR simulator.	77	Objective measurement focussing on measuring proficiency.	Specific to the Mimic simulator. Specific to VR. No correlation with operative outcomes.	[54]
Robotic Hysterectomy Assessment Score (RHAS)	Deconstructs RHAS into 6 key domains- each assessed using a 5-point Likert scale.	52	Construct validity. Procedural specific. Developed using video recordings of RHAS.	Assessed 7 novice surgeons therefore construct validity is questionable.	[83]
Clinically Relevant Objective Measures (CROMS)	Clinically proficient metrics pertinent to clinical outcomes.	43	Demonstrates construct validity. Objective measures relative to a complex procedure.	Challenging to determine which are clinically relevant metrics. Limited studies assessing validity.	[19]
Pupillary measures	Measures of pupil dilatation and diameter in response to task demands.	43	Highly objective measure of cognitive function and stress response.	Not widely available- dependent on using eye tracking systems.	[49, 84, 85]

Psychomotor vigilance task	Measures reaction time to tasks presented as short random intervals	40	Highly sensitive for measuring deficits in sustained attention.	Uncertainty with regards relevance to RAS performance.	[50]
Wisconsin Card Sorting Test (WCST)	Gold standard measure of executive function.	40	Validated tool for measuring executive function.	Uncertainty with regards relevance to RAS performance.	[50]
Proficiency Based Progression (PBP)	Deconstruction of a surgical procedure to operational defined performance metrics.	38	Demonstrated to reduce errors by 40% Demonstrates construct validity.	Time consuming to develop the metric. Requires an expert group to develop the metrics- difficult to define. Requires dedicated time for the development of each procedural PBP	[17, 44]
Camera Metrics	Comprised of 3 camera metrics: (1) camera movement frequency, (2) camera measurement duration and (3) camera movement interval.	39	Demonstrates construct validity. Correlation with efficiency metrics across VR simulation exercises.	Limited evidence for use.	[48]
Global Rating Score (GRS)	Sum of all individual modified OSATs score.	34	Easy to use and implement. Demonstrates construct validity.	Dependent on Likert scales therefore subject to variation amongst users.	[64, 69, 86]
Operative Room Time (ORT)	Measures time from skin incision to closure	28	Easy to measure and implement. Applicable to all procedures	Non-specific measure of performance	[87, 88]
Robotic Skills Assessment Score (RSA) score	Developed using the Fundamentals of Robotic Surgery curriculum metrics. Composed of safety in operative field.	27	Able to validate and integrate simulated scores into a scoring system to assess performance.	Tested only on RoSS VR simulator and not in clinical settings. Tested on a small study sample size.	[89]
Prostatectomy assessment and competency evaluation (PACE)	Deconstructed RARP videos into 7 key domains. Proficiency within each domain assessed.	26	Modular training with specific descriptions of scores assigned to each stage	Reliant on pre-recorded videos; not real time surgery.	[39]

				Has not been widely tested	
aOSATs	Modified version of OSATs used to assess surgical assistance performance.	26	Unique global measure specific to assistant's performance	Reliant upon Likert scales. Not widely tested.	[20]
Objective Structured Assessment Tools (OSATs)	5-point Likert scale assessing global measures of performance; (1) respect for tissues (2) time & motion (3) instrument handling (4) knowledge of instruments (5) use of assistants (6) flow of operation and forward planning (7) knowledge of specific procedure.	22	Easy to use and widely implementable.	Time consuming to complete. Dependent on Likert scales therefore subject to variation amongst users.	[90]
Numeric psychomotor test score (NPMTS)	32-point psychomotor test score.	21	Validated metric scoring system previously used for FRS curriculum.	Time consuming to complete	[91]
Robotic Assessment of Radical Prostatectomy (RARP)	RARP deconstructed into three phases: (1) preparation of operative field (2) dissection of bladder and (3) prostate anastomosis	15	Non-time consuming to complete. Provides a modular training pathway for RARP. Demonstrates construct validity	No assessment of non-technical skills.	[40]
Assessment of Robotic Console Skills (ARCS)	Assesses 5 skill domains using a 5 point Likert scale; (1) dexterity with multiple wristed instruments (2) optimising field of view (3) instrument visualisation (4)optimising master manipulator workspace (5) force sensitivity and control (6) basic energy pedal skills.	15	Demonstrates constructs validity amongst 5 out of 6 domains. Console agnostic.	Only assessed on an animal model. Requires views of the surgeons' hands, feet and operative field therefore 3 separate video recordings needed.	[92]
Structured assessment of robotic microsurgical skills	Uses 3 parametres to assess microsurgical skills (1) dexterity (2) visuospatial ability (3) operational flow. Additional 5	10	Metric designed specifically for robotic microsurgical skills.	Dependent on Likert scales therefore subject to variation amongst users. Lacks construct validity.	[93]

	parametres to assess robotic skills.				
Operative Room Index (ORI)	Significant kinematic variables from DV logger combined	2	Used to model the learning curve for novice HPB surgeons.	Small sample size. Uncertainty regarding wider use.	[94]
Generic Error Rating Tool (GERT)	Records number of errors performed by the operating surgeon and bed side assistant throughout the procedure.	1	Global assessment of errors of both the surgeon and assistant.	Some errors and events captured may be clinically insignificant- a broad definition of errors including near misses are included in the assessment.	[30]
Cystectomy Assessment and surgical evaluation (CASE)	Critical steps of robotic assisted cystectomy deconstructed into 8 key domains each assessed by a 5-point Likert scale.	N/a	Procedural specific Developed using real surgical performance.	Designed specifically for male cystectomy. Lacked statistical significance in assessing construct validity	[41]

A description of the studies included in the review are listed in Table 3. The majority of the included studies were observational (n=67 93%). Only 5 (7%) studies involved a randomised controlled trial (RCT).

On reviewing the modalities used for training amongst the included studies, simulation was the most utilised (n=50, 69%). VR simulation was the most common modality of simulation, (n=24, 44%) followed by dry lab simulation using bench top models (n=20, 40%).

Quality assessment of the studies included.

Using the Newcastle Ottawa Score, 34 studies (47%) were assessed as fair quality with an average risk of bias. 29 studies (40%) studies were assessed as good quality, n=5 (7%) as poor quality. Only 5 (7%) studies using randomised controlled trials were included. Using the Cochrane Risk Bias tool to assess the included RCT, s n=3 (60%) risk of bias was unclear, n=1 (20%) was high risk for bias and n=1 (20%) had a low risk of bias.

Table 3: Description of included studies.

Authors	Year of Publication	Study Type	No. of Participants	Skill assessed	Performance measure	Modality of Training	Quality Assessment (Newcastle Ottawa for Observational studies and Cochrane Risk Bias tool for RCTs)
Alrasheed T et al.	2022	Observational	10 trainees (varying levels)	Micro-surgical anastomosis	Structured assessment of robotic microsurgery skills.	Dry lab Simulation	Good
Hung A. et al.	2022	Observational	22: 7 trainees and 15 experts.	Suturing	APMs	VR Simulated	Fair
Trinh et al.	2022	Observational	23 surgeons of varying expertise.	Vesicourethral Anastomosis (VUA) & prostatectomy	APMs and RACE	Video-analysis	Good
Gomez-Ruis M.et al.	2022	Observational	14 participants: 9 senior surgeons, 5 novices.	Robotic Low Anterior Resection (RA-LAR)	GEARS, PBP	Video analysis	Good
.Oğul et al.	2022	Observational	12 trainees.	Basic	APMs	VR simulation	Poor
Martin et al.	2021	Observational	20 novice trainees.	Basic skills	GEARS	VR simulation	Fair
Simmonds et al.	2021	Observational	77 students (residents to practicing surgeons)	Basic skills	APMs, MScore Proficiency Index (MPI©)	VR Simulation	Fair
Wu et al.	2021	Observational	7 novice trainees.	Basic Skills	APMs, Cognitive and Behavioural metrics including gaze entropy and NASA TLX.	VR Simulation	Good
Cowan et al.	2021	Observational	17 participants: 6 experts, 11 trainees.	VUA	APMs, pupillary data	Simulation- Dry lab and VR	Good
Ghodoussipour et al.	2021	Observational	27 participants: 10 experts, 17 trainees	Partial nephrectomy	APMs	Video recordings of RAPN	Good
Chen A et al.	2021	Observational	N/a	Prostatectomy	APMs	Video recordings of	Fair

						RAPN	
Chow A et al.	2021	Observational	12 residents.	Partial Nephrectomy	GEARS	Simulation- Animal model	Good
Ghazi A et al.	2021	Observational	43 participants: 27 novices, 16 Experts.	Nephrectomy	GEARS, Clinically Relevant Objective Metrics of Simulators) CROMS	Simulation	Fair
Mottrie A et al.	2021	Observational	12 novices, 12 experts.	Robotic Assisted Radical Prostatectomy (RARP)	PBP	Video recordings of RARP	Good
Yu N et al.	2021	Observational	14 staff urologists, 22 bed-side assistants with variable experience.	RARP	GEARS, aOSATs.	Live RARP procedures	Fair
Tou S et al.	2020	Observational	Metrics group-4 including 3 robotic colorectal surgeons and one behavioural scientist. Delphi group-18 colorectal surgeons.	Robotic Low Anterior resection (RA-LAR)	PBP	Video recordings of LAR	Good
Lyman W et al.	2020	Observational	2 hepatobiliary fellows.	Robotic hepaticojejunostomy	Operative Robotic Index (ORI)	Dry lab Simulation	Fair
Møller SG et al.	2020	RCT	22 novice trainees.	Suturing	OSATs	Simulation	Unclear
Rice MK et al.	2020	Observational	28 surgical trainees -varying experience.	Robotic Pancreaticoduodenectomy	Operating Room Time (ORT)	Live operating	Good
Lau E et al.	2020	RCT	40 surgical trainees: 16 senior residents and 14 junior residents.	Basic Skills	NASA- Task load index- subjective workload Wisconsin Card Sorting Test (WCST)- executive cognitive function Psychomotor Vigilance Task (PVT)- concentration	Animal model simulation	Unclear
Brown KC et al	2020	Observational	Surgical trainees (Trainees, experts and training specialists defined as non-surgeon expert users who were experienced in the trained exercises).	Basic skills	APMs	Animal model simulation	Good
Nguyen JH et al	2020	Observational	26 participants-stratified by surgical experience.	Basis skills	APMs, Task-evoked pupillary response (TEPR)	Dry lab simulation	Fair
Dilley J et al.	2020	Observational	21 surgical trainees-novice and Experts.	Basic skills	GEARS, Numeric psychomotor test score (NPMTS)	Animal simulation model	Good

Lefor A et al.	2020	Observational	8 surgical trainees: 4 novices, 2 intermediates and 2 experts.	Basic skills	APMs	VR simulation	Good
Timberlake M et al.	2020	Observational	25 participants-urology residents 5 faculty surgeons, 6 fellows, 14 residents	Pyeloplasty	GEARS	Dry lab simulation- 3D models	Good
Ebbing J et al	2020	Observational	51 participants:18 experts, 16 intermediates, 17 novices	Radical Prostatectomy	APMs	VR Simulation	Fair
Sánchez R et al.	2019	Observational	15 participants divided into novice, intermediate and expert.	Suturing	GEARS	Dry lab simulation	Fair
Witthaus M et al.	2020	Observational	14 participants: 9 novices, 5 experts.	Radical prostatectomy	GEARS, RACE	Dry lab simulation	Good
Khan H et al.	2019	Observational	6 surgical trainees.	Vesicourethral anastomosis (VUA) in RARP.	RACE	Live RARP	Fair
Peng W et al.	2018	Observational	14 participants: 4 experts, 10 novices.	Basic skills	APMs	VR Simulation	Good
Liu M et al	2018	Observational	15 surgical trainees- divided into senior, intermediate and novice	Basic skills	Assessment of Robotic Console Skills (ARCS)	Animal model simulation	Fair
Knab L et al.	2018	Observational	28 surgical trainees	Pancreatico-duodenectomy	APMs	Training pathway for pancreatico-duodenectomy	Fair
Hoogenes et al.	2018	RCT	39 surgical trainees- 23 juniors and 16 seniors.	Vesicourethral anastomosis (VUA)	GEARS, RACE (Robotic Anastomosis Competency Evaluation) score	Dry lab simulation- 3D printed model	Low
Dubin A et al.	2018	Observational	65 surgical trainees.	Basic skills	APMs, GEARS	VR Simulation	Good
Zia A et al.	2018	Observational	8 surgical trainees of varying experience.	Suturing	APMs, modified OSATs and Global Rating Score	Dry lab simulation	Fair
Wang Z et al.	2018	Observational	10 surgical trainees (divided as novice, intermediate and expert).	Basic skills	APMs	Dry lab simulation	Good
Guni et al.	2018	Observational	39 novices.	Suturing	GEARS	Dry lab Simulation of a VUA	Good

Watkinson et al.	2018	Observational	123: 84 novice,6 beginners intermediates, 9 advanced intermediates and 4 experts.	Basic Skills	APMs	VR Simulation	Good
Altok et al.	2018	Observational	100 participants: 94- trainees (43 fellows, 51 residents) and 6 faculty surgeons.	RARP	Time to completion	Video recordings of RARP	Fair
Hussein A et al.	2018	Observational	No of trainees and trainers not stipulated	Robotic assisted radical cystectomy	CASE score	Video recording of RARC	Fair
Shim J et al.	2018	Observational	3 trainees.	VUA anastomosis	Mean completion times.	VR Simulation	Fair
Chen J et al.	2018	Observational	18 participants: 9 novices and 9 experts.	VUA	APMs	Video recordings of RARP	Good
Hung A et al.	2018	Observational	9 faculty surgeons.	RARP	APMs	Recorded RARP	Poor
Dubin A et al	2017	RCT	65 surgical trainees.	Basic skills	APMs, GEARS	VR Simulation	Unclear
Sessa A et al.	2018	Observational	21 surgical trainees– 12 beginners, 9 experts.	Basis skills	NASA- Task load index	VR Simulation	Fair
Mills J et al.	2017	Observational	10 attending robotic surgeons.	Basic skills	GEARS	DaVinci simulator and intraoperative video clips.	Fair
Fard M et al.	2017	Observational	8 surgical trainees divided into seniors and novices.	Basic skills	APMs, Global Rating Score	Dry lab Simulation	Fair
Lee G et al.	2017	Observational	32 surgical trainees.	Basic skills	APMs, NASA-TLX	VR Simulation	Poor
Raison N et al.	2016	Observational	223 participants ranging from novice to expert surgeons.	Basic skills	APMs	VR Simulation	Good
Hussein A et al.	2017	Observational	26 participants: 23 experts, 3 fellows.	Prostatectomy	Prostatectomy Assessment and Competence Evaluation (PACE)	Video recording of RARP	Fair
Hung A et al.	2018	Observational	20 trainees: 10 experts and 10 trainees.	RARP	APMs, GEARS	Video recordings of RARP	Good
Goldenberg M et al.	2017	Observational	1 consultant urologist.	RARP	GEARS, GERT	Video recordings of RARP	Poor
Hung A et al.	2017	Observational	21 trainees (11 residents and 10 fellows)	Prostatectomy and nephrectomy	GEARS, proficiency score	Video recordings of RARP and RPN	Fair

Jarc AM et al.	2017	Observational	39 trainees- divided into groups by RAS experience (new, intermediate and novice)	Basic skills	APMs, camera metrics	VR Simulation	Fair
Frederick P et al.	2016	Observational	52- 25 expert surgeons, 20 advanced beginners, 7 novice trainees.	Robotic Hysterectomy	Robotic Hysterectomy Assessment Score (RHAS)	Video assessment of live procedures	Good
Vedula S. et al.	2016	Observational	18 participants: 4 consultant surgeons (experts), 14 (novice) trainees.	Suturing	APMs, Global Rating Score	Dry lab simulation	Fair
Siddiqui N et al.	2016	Observational	46 participants: 34 novice and 22 senior trainees.	Basic Skills	R-OSATs	Dry lab simulation	Fair
Aghazadeh A et al.	2016	Observational	21 participants: 17 trainees, 4 experts.	RARP	APMs, GEARS	VR Simulation- APMs. Live operating- GEARs	Good
Lovegrove et al.	2016	Observational	15 fellows and trainees.	RARP	RARP-score	Video recordings Of live RARP	Fair
White L et al.	2015	Observational	49 surgeons.	Basic skills	GEARS	Dry lab Simulation	Fair
Tanaka A et al.	2015	Observational	105: novice-37, intermediate-31, and expert-37	Basic skills	APMs	VR Simulation	Poor
Aghazadeh M et al.	2015	Observational	47 surgical trainees: experts- 9, Intermediates- 14, novices- 24	Suturing	GEARS	Animal model Simulation	Fair
Whitehurst S et al.	2015	RCT	23 participants: 20 novices and 3 experts.	Cystostomy closure on an animal model	APMs, GEARS	Animal model Simulation	High
Raza S et al.	2015	Observational	28 participants: 8 experts, 10 advanced beginners, 10 novice surgeons.	VUA	RACE	Dry lab simulation	Good
Siddiqui N. et al.	2014	Observational	105: 83 residents, 9 fellows and 13 faculty surgical trainees of varying levels	Basic skills	R-OSATs	Dry lab Simulation	Good
Chen C et al.	2014	Observational	476 participants	Suturing	GEARs, C-SATs	Dry lab simulation	Fair
Chowriappa A et al.	2013	Observational	27 participants: 15 novice and 12 expert surgeons	Basic robotic skills	RSA-score	Simulation-VR	Fair

Perrenot C et al	2012	Observational	75- divided by robotic experience	Basic skills	APMs	VR Simulation	Fair
Goh A et al.	2012	Observational	29 participants: 25 trainees, 4 experts.	Radical prostatectomy	GEARS	Live operating RAS	Fair
Kumar R et al.	2012	Observational	8 participants:6 trainees, 2 expert surgeons.	Basic skills	APMs	Dry lab simulation	Good
McDonough P et al.	2011	Observational	20 participants: 10 trainees(novice), 10 consultants(experts)	Basic skills	APMs	Dry lab simulation	Fair
Judgkins T et al.	2008	Observational	1 consultant surgeon-expert.	Suturing	APMs	Live operating-human cases	Fair

Discussion

We have systematically reviewed the metrics currently used for assessing technical performance in RAS. This is the first review to assess metrics used to evaluate technical performance across all surgical specialties. Twenty-seven metrics were identified and categorised into four groups; global skills assessment; procedural-based assessment; task based and cognitive assessment. This review has highlighted the wide diversity in metrics available but emphasises the predominant use of global assessments owing to their broad application.

Global-based assessment

GEARS was the first global assessment tool designed specifically for RAS[37] and has the advantage of being widely implementable across RAS procedures. GEARS demonstrated construct validity and was later found to demonstrate a significant relationship between simulated robotic performance and robotic clinical performance[32].

This review found GEARS has superseded OSATs [95] due to its granular assessment of RAS skills. Adaptations of OSATs have been developed which are specific to RAS surgery including aOSATs and ROSATs however both were confined to single studies[20, 80]. The Global Rating Score (GRS) [64, 69, 72] is a mean of the calculated OSATs scores and therefore confers the same advantages and disadvantages of OSATs (Table 3).

Assessment of Robotic Console Skills (ARCS) relies upon a similar assessment to GEARS and has the advantage of being console agnostic, however, within the review it was only tested on an animal model and a limited number of participants (n=15).

Other subjective metrics were found in the review including the operative room index (ORI) [94] and the Global Error Rating Tool (GERT) [30], however both were used in small studies (<5 participants).

CROMs were the most frequently used however they were only assessed in a simulated environment[38]. Operative Room Time (ORT) were used in two studies [87, 88], but the tool is a non-specific measure of performance and is therefore unable to demonstrate construct validity. The MScore Proficiency Index score (MPI©) was tested only in one paper[54] and is limited to the Mimic VR simulator therefore, it is unclear if the metric can be applied in the clinical setting.

Procedural-Based Assessments

The RARP score[40] and the Prostatectomy Assessment and Competency Evaluation (PACE) [39] both deconstruct a RARP procedure into key stages. Both scores had content validity, however, although there were fewer participants in the RARP score study it was used across a larger number of RARP videos (n=428). Both scores provide a modular training pathway for RARP, however, neither assessed non-technical skills such as communication skills.

The Cystectomy Assessment and Surgical Evaluation score (CASE) was specific to a robotic cystectomy. The metric however is limited to male cystectomies.

The Robotic Hysterectomy Assessment Score[83] was specific only to a robotic hysterectomy. It demonstrated construct validity; however, the number of novice surgeons were small therefore the validity is questionable. Crowd-Sourced Assessment Tools (C-SATs)[36] is a unique procedural metric which can be applied to any procedure. It is a quick assessment tool to use but is reliant upon GEARS assessment.

Proficiency Based Progression (PBP) was the only objective measure [42-44] and the most widely used procedural-specific metric. The PBP methodology focuses on training towards proficiency and reduction of technical error rates. It was used in several studies and demonstrated construct validity [17, 44]. The PBP methodology focusses on proficiency training and has demonstrated a reduction in errors by up to 40% [96]. However, the process to develop the metrics is lengthy and is reliant upon a Delphi process involving an expert group. This will impact implementation on a wider scale and therefore questions its feasibility in curriculum development.

Task-Based assessments

RACE was the most prevalent subjective task-based assessment used. It demonstrated construct validity, however was specific to measuring one specific urological task only and therefore is not widely applicable across RAS assessments.

aOSATs[20] specifically assesses the assistant's performance during a procedure, however, it was only assessed in one study.

APMs were the most common objective task-based metric. APMs summarise key elements of intraoperative surgical behaviour, particularly geometric and time-dependent variables of instrument and surgical console control, providing a complex and comprehensive data record for each operation. Early studies have shown that APM may predict progression in surgical training, and also predict when expertise is gained [97]. APMs have also been able to differentiate between novice and expert surgeon (i.e., demonstrate construct validity) in live prostate surgery and video recordings of a vesicourethral anastomosis (VUA)[98]. In addition, APMs can predict important clinical outcomes including length of stay and urinary continence following RARP [56, 57] [68]. In the simulation setting, APMs can be recorded in VR scenarios thereby aiding a trainee's learning and providing valuable feedback against bench markers.

However, at present their use is largely limited to measuring task-based measures. Furthermore, the APMs used in clinical settings were inclusive to Intuitive Surgical only and therefore it is unclear if the APMs are platform agnostic which questions their wider application.

Cognitive Metrics

NASA-TLX was the only subjective metric found which specifically tailors to the surgical environment. Pupillary metrics were the most prevalent object cognitive metric however they are dependent upon prohibitively expensive eye tracking glasses which will be challenging to implement alongside current training methods.

The remaining cognitive measures, Numeric Psychomotor Test Score (NPMTS) and Wisconsin Card Sorting Test (WCST) were limited to single studies. Their implications in a wider setting are therefore unclear. Despite the limited use of cognitive measures in this review, they were used and were able to assess the importance in assessing a of a surgeon's mental workload on performance[50]. These measures could possibly help a trainee self-monitor their progression.

It is clear from the variation in metrics defined and their various uses in different settings and surgical subspecialities, an ideal metric is needed to standardise RAS assessment thereby widening its applications across procedures. This will facilitate both surgical trainers and trainees with the accuracy and interpretation of RAS assessment.

The "Ideal" Metric for RAS technical assessment

There is increasing evidence a surgeon's technical skill can affect postoperative outcomes [57]. The method chosen to train surgeons and provide feedback has previously been shown to result in efficient surgical training[58]. On developing the ideal metric for assessment, it should have components of global, procedural, task-based, and cognitive performance measures to provide an accurate assessment of the trainee's performance.

An ideal metric would be difficult to define and even harder to develop. It could utilise the versatility of automated performance metrics (APMs), allowing wide application and accuracy in assessing and recording performance measures. However, it could also use the principles of PBP to assess proficiency of an index procedure at a granular level. PBP has demonstrated reliability however lacks feasibility for wider implementation secondary to the lengthy process to develop the metric. As APMs advance, they may be used to develop proficiency-based metrics similar to PBPs thereby having the ability to assess procedural based skills at a granular level. Automation will inevitably accelerate the process of development and will facilitate with wider applications across surgical specialties. This novel metric which is an objective measure, will allow for efficient and reliable measures of technical performance in RAS.

This review has identified limitations in the development of metric assessment for technical performance which are discussed below with suggestions for improvement.

Standardisation of validity assessment

To assess the validity of the metric included in this review, the ability to demonstrate construct validity was evaluated. Construct validity is described as the extent to which measurements used can actually test the hypothesis or theory it is measuring[99]. Construct validity is commonly used within robotic literature as a comparison between experienced and novice surgeons, however, there are variations amongst the studies regarding the definitions of novice and expert surgeons. Novices varied in their experience of RAS from none to minimal to some. [59, 91]. The discrepancy in definition results in the construct validity of the metric being questioned. A standardised definition of both the novice and expert with defined levels of technical experience is required to generate a valid assessment of the construct validity of the metric.

Assessment of non-technical skills for surgeons (NOTSS)

NOTSS assessment was not included in this review, however, they are an essential component of RAS training. The recently published report by the Royal College of Surgeons Edinburgh "*Development of new robotic surgical services- A guide to good practice"* provides generic recommendations covering areas of training and clinical governance to all surgical specialties[100]. The report recommends training should focus on the non-technical skills of the robotic surgical team's performance as the dynamics in the surgical team are different from conventional surgery. A validated non-technical skills framework should be employed to standardise training in interpersonal communication, teamwork and situational awareness thereby safeguarding communication in robotic-assisted cases.

Conclusion

The end goal of the assessment metric should be to measure and assist in achievement of proficiency in RAS procedural training. On reviewing the metrics that are currently available we suggest that there is a need to develop a robust method that is able to accurately assess technical performance in RAS. We believe a novel automated metric which incorporates components of the categorised metrics already identified should be considered for further development. By incorporating the benefits shown by the categorised metrics, it will provide a granular assessment of a surgeon's skills performance from task-based assessment to cognitive. Automated metrics have shown promise in their wide use and ease to assess objective performance. However, further work is needed to assess their use in measuring clinical outcome.

References

1. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. 2013;369(15):1434-42.
2. Curtis NJ, Foster JD, Miskovic D, Brown CSB, Hewett PJ, Abbott S, et al. Association of Surgical Skill Assessment With Clinical Outcomes in Cancer Surgery. JAMA Surg. 2020.
3. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat A-HS, Dellinger EP, et al. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. New England Journal of Medicine. 2009;360(5):491-9.
4. Jung JJ, Juni P, Lebovic G, Grantcharov T. First-year Analysis of the Operating Room Black Box Study. Ann Surg. 2020;271(1):122-7.
5. Garbens A, Goldenberg M, Wallis CJD, Tricco A, Grantcharov TP. The cost of intraoperative adverse events in abdominal and pelvic surgery: A systematic review. Am J Surg. 2018;215(1):163-70.
6. 2023 IArAahwacCi-s-iAtF. [
7. Bravi CA, Paciotti M, Balestrazzi E, Piro A, Piramide F, Peraire M, et al. Outcomes of Robot-assisted Radical Prostatectomy with the Hugo RAS Surgical System: Initial Experience at a High-volume Robotic Center. European Urology Focus. 2023.
8. Butterworth J, Sadry M, Julian D, Haig F. Assessment of the training program for Versius, a new innovative robotic system for use in minimal access surgery. BMJ Surg Interv Health Technol. 2021;3(1):e000057.
9. McCrum ML, Valmont T, Price RR. Developing a Surgical Quality Improvement Program for Resource-Limited Settings. JAMA Surg. 2020.
10. F D, BD K. Robotic surgery: training, competence assessment and credentialing. The Bulletin of the Royal College of Surgeons of England. 2020;102(7):302-6.
11. Fleming CA, Ali O, Clements JM, Hirniak J, King M, Mohan HM, et al. Pan-specialty access to robotic surgery in surgical training. British Journal of Surgery. 2021;108(7):e245-e6.
12. Challacombe B, Ahmed K, Soomro N, Dasgupta P, Khan M, Cross W, et al. British Association of Urological Surgeons (BAUS) Robotic Surgery Curriculum: Guideline for Training. London: BAUS Business. 2015.
13. Burke JR, Fleming CA, King M, El-Sayed C, Bolton WS, Munsch C, et al. Utilising an accelerated Delphi process to develop consensus on the requirement and components of a pre-procedural core robotic surgery curriculum. J Robot Surg. 2023:1-13.
14. Younes MM, Larkins K, To G, Burke G, Heriot A, Warrier S, et al. What are clinically relevant performance metrics in robotic surgery? A systematic review of the literature. Journal of Robotic Surgery. 2023;17(2):335-50.
15. Chen J, Cheng N, Cacciamani G, Oh P, Lin-Brande M, Remulla D, et al. Objective Assessment of Robotic Surgical Technical Skill: A Systematic Review. Journal of Urology. 2019;201(3):461-9.
16. Martin JR, Stefanidis D, Dorin RP, Goh AC, Satava RM, Levy JS. Demonstrating the effectiveness of the fundamentals of robotic surgery (FRS) curriculum on the RobotiX Mentor Virtual Reality Simulation Platform. Journal of Robotic Surgery. 2021;15(2):187-93.
17. Gomez Ruiz M, Tou S, Gallagher AG, Cagigas Fernandez C, Cristobal Poch L, Matzel KE. Intraoperative robotic-assisted low anterior rectal resection performance assessment using procedure-specific binary metrics and a global rating scale. BJS open. 2022;6(3).

18. Chow AK, Wong R, Monda S, Bhatt R, Sands KG, Vetter J, et al. Ex Vivo Porcine Model for Robot-Assisted Partial Nephrectomy Simulation at a High-Volume Tertiary Center: Resident Perception and Validation Assessment Using the Global Evaluative Assessment of Robotic Skills Tool. J Endourol. 2021;35(6):878-84.
19. Ghazi A, Melnyk R, Hung AJ, Collins J, Ertefaie A, Saba P, et al. Multi-institutional validation of a perfused robot-assisted partial nephrectomy procedural simulation platform utilizing clinically relevant objective metrics of simulators (CROMS). BJU Int. 2021;127(6):645-53.
20. Yu N, Saadat H, Finelli A, Lee JY, Singal RK, Grantcharov TP, et al. Quantifying the "Assistant Effect" in Robotic-Assisted Radical Prostatectomy (RARP): Measures of Technical Performance. J Surg Res. 2021;260:307-14.
21. Dilley J, Singh H, Pratt P, Omar I, Darzi A, Mayer E. Visual behaviour in robotic surgery-Demonstrating the validity of the simulated environment. Int J Med Robot. 2020;16(2):e2075.
22. Timberlake MD, Garbens A, Schlomer BJ, Kavoussi NL, Kern AJM, Peters CA, et al. Design and validation of a low-cost, high-fidelity model for robotic pyeloplasty simulation training. Journal of Pediatric Urology. 2020;16(3):332-9.
23. Sanchez R, Rodriguez O, Rosciano J, Vegas L, Bond V, Rojas A, et al. Robotic surgery training: construct validity of Global Evaluative Assessment of Robotic Skills (GEARS). Journal of robotic surgery. 2016;10(3):227-31.
24. Witthaus MW, Farooq S, Melnyk R, Campbell T, Saba P, Mathews E, et al. Incorporation and validation of clinically relevant performance metrics of simulation (CRPMS) into a novel full-immersion simulation platform for nerve-sparing robot-assisted radical prostatectomy (NS-RARP) utilizing three-dimensional printing and hydrogel casting technology. BJU Int. 2020;125(2):322-32.
25. Hoogenes J, Wong N, Al-Harbi B, Kim KS, Vij S, Bolognone E, et al. A Randomized Comparison of 2 Robotic Virtual Reality Simulators and Evaluation of Trainees' Skills Transfer to a Simulated Robotic Urethrovesical Anastomosis Task. Urology. 2018;111:110-5.
26. 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. J Minim Invasive Gynecol. 2017;24(7):1184-9.
27. Guni A, Raison N, Challacombe B, Khan S, Dasgupta P, Ahmed K. Development of a technical checklist for the assessment of suturing in robotic surgery. Surg Endosc. 2018;32(11):4402-7.
28. Mills JT, Hougen HY, Bitner D, Krupski TL, Schenkman NS. Does Robotic Surgical Simulator Performance Correlate With Surgical Skill? Journal of Surgical Education. 2017;74(6):1052-6.
29. Hung AJ, Chen J, Jarc A, Hatcher D, Djaladat H, Gill IS. Development and Validation of Objective Performance Metrics for Robot-Assisted Radical Prostatectomy: A Pilot Study. The Journal of Urology. 2018;199(1):296-304.
30. Goldenberg MG, Goldenberg L, Grantcharov TP. Surgeon Performance Predicts Early Continence After Robot-Assisted Radical Prostatectomy. J Endourol. 2017;31(9):858-63.
31. Hung AJ, Bottyan T, Clifford TG, Serang S, Nakhoda ZK, Shah SH, et al. Structured learning for robotic surgery utilizing a proficiency score: a pilot study. World J Urol. 2017;35(1):27-34.

32. 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):475-81.
33. Aghazadeh MA, Jayaratna IS, Hung AJ, Pan MM, Desai MM, Gill IS, et al. External validation of Global Evaluative Assessment of Robotic Skills (GEARS). Surg Endosc. 2015;29(11):3261-6.
34. 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):1295-301.
35. Whitehurst SV, Lockrow EG, Lendvay TS, Propst AM, Dunlow SG, Rosemeyer CJ, 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):483-8.
36. Chen C, White L, Kowalewski T, Aggarwal R, Lintott C, Comstock B, et al. Crowd-Sourced Assessment of Technical Skills: a novel method to evaluate surgical performance. J Surg Res. 2014;187(1):65-71.
37. 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. J Urol. 2012;187(1):247-52.
38. Ghazi A, Melnyk R, Hung AJ, Collins J, Ertefaie A, Saba P, et al. Multi-institutional validation of a perfused robot-assisted partial nephrectomy procedural simulation platform utilizing clinically relevant objective metrics of simulators (CROMS). BJU International. 2021;127(6):645-53.
39. Hussein AA, Ghani KR, Peabody J, Sarle R, Abaza R, Eun D, et al. Development and Validation of an Objective Scoring Tool for Robot-Assisted Radical Prostatectomy: Prostatectomy Assessment and Competency Evaluation. Journal of Urology. 2017;197(5):1237-44.
40. Lovegrove C, Novara G, Mottrie A, Guru KA, Brown M, Challacombe B, 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. 2016;69(3):526-35.
41. Hussein AA, Sexton KJ, May PR, Meng MV, Hosseini A, Eun DD, et al. Development and validation of surgical training tool: cystectomy assessment and surgical evaluation (CASE) for robot-assisted radical cystectomy for men. Surgical Endoscopy. 2018;32(11):4458-64.
42. Gómez Ruiz M, Tou S, Gallagher AG, Cagigas Fernández C, Cristobal Poch L, Matzel KE. Intraoperative robotic-assisted low anterior rectal resection performance assessment using procedure-specific binary metrics and a global rating scale. BJS Open. 2022;6(3).
43. Mottrie A, Mazzone E, Wiklund P, Graefen M, Collins JW, De Groote R, et al. Objective assessment of intraoperative skills for robot-assisted radical prostatectomy (RARP): results from the ERUS Scientific and Educational Working Groups Metrics Initiative. BJU Int. 2021;128(1):103-11.
44. Tou S, Gómez Ruiz M, Gallagher AG, Matzel KE. European expert consensus on a structured approach to training robotic-assisted low anterior resection using performance metrics. Colorectal Dis. 2020;22(12):2232-42.
45. Raza SJ, Field E, Jay C, Eun D, Fumo M, Hu JC, 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.

46. Trinh L, Mingo S, Vanstrum EB, Sanford DI, Aastha, Ma R, et al. Survival Analysis Using Surgeon Skill Metrics and Patient Factors to Predict Urinary Continence Recovery After Robot-assisted Radical Prostatectomy. Eur Urol Focus. 2022;8(2):623-30.
47. Khan H, Kozlowski JD, Hussein AA, Sharif M, Ahmed Y, May P, et al. Use of Robotic Anastomosis Competency Evaluation (RACE) tool for assessment of surgical competency during urethrovesical anastomosis. Canadian Urological Association Journal. 2019;13(1):E10-E6.
48. Jarc AM, Curet MJ. Viewpoint matters: objective performance metrics for surgeon endoscope control during robot-assisted surgery. Surg Endosc. 2017;31(3):1192-202.
49. Wu C, Cha J, Sulek J, Sundaram CP, Wachs J, Proctor RW, et al. Sensor-based indicators of performance changes between sessions during robotic surgery training. Appl Ergon. 2021;90:103251.
50. Lau E, Alkhamesi NA, Schlachta CM. Impact of robotic assistance on mental workload and cognitive performance of surgical trainees performing a complex minimally invasive suturing task. Surgical Endoscopy. 2020;34(6):2551-9.
51. Sessa L, Perrenot C, Xu S, Hubert J, Bresler L, Brunaud L, et al. Face and content validity of XperienceTM Team Trainer: bed-side assistant training simulator for robotic surgery. Updates in surgery. 2018;70(1):113-9.
52. Hung AJ, Rambhatla S, Sanford DI, Pachauri N, Vanstrum E, Nguyen JH, et al. Road to automating robotic suturing skills assessment: Battling mislabeling of the ground truth. Surgery. 2022;171(4):915-9.
53. Oğul BB, Gilgien M, Özdemir S. Ranking surgical skills using an attention-enhanced Siamese network with piecewise aggregated kinematic data. Int J Comput Assist Radiol Surg. 2022;17(6):1039-48.
54. Simmonds C, Brentnall M, Lenihan J. Evaluation of a novel universal robotic surgery virtual reality simulation proficiency index that will allow comparisons of users across any virtual reality simulation curriculum. Surg Endosc. 2021;35(10):5867-75.
55. Cowan A, Chen J, Mingo S, Reddy SS, Ma R, Marshall S, et al. Virtual Reality vs Dry Laboratory Models: Comparing Automated Performance Metrics and Cognitive Workload During Robotic Simulation Training. J Endourol. 2021;35(10):1571-6.
56. Ghodoussipour S, Reddy SS, Ma R, Hwang DH, Nguyen J, Hung AJ. An Objective Assessment of Performance during Robotic Partial Nephrectomy: Validation and Correlation of Automated Performance Metrics with Intraoperative Outcomes. Journal of Urology. 2021;205(5):1294-302.
57. Chen A, Ghodoussipour S, Titus MB, Nguyen JH, Chen J, Ma R, et al. Comparison of clinical outcomes and automated performance metrics in robot-assisted radical prostatectomy with and without trainee involvement. World J Urol. 2020;38(7):1615-21.
58. Brown KC, Bhattacharyya KD, Kulason S, Zia A, Jarc A. How to bring surgery to the next level: Interpretable skills assessment in robotic-assisted surgery. Visceral Medicine. 2020;36(6):463-70.
59. Nguyen JH, Chen J, Marshall SP, Ghodoussipour S, Chen A, Gill IS, et al. Using objective robotic automated performance metrics and task-evoked pupillary response to distinguish surgeon expertise. World journal of urology. 2020;38(7):1599-605.
60. Lefor AK, Harada K, Dosis A, Mitsuishi M. Motion analysis of the JHU-ISI Gesture and Skill Assessment Working Set using Robotics Video and Motion Assessment Software. Int J Comput Assist Radiol Surg. 2020;15(12):2017-25.

61. Ebbing J, Wiklund PN, Akre O, Carlsson S, Olsson MJ, Höijer J, et al. Development and validation of non-guided bladder-neck and neurovascular-bundle dissection modules of the RobotiX-Mentor® full-procedure robotic-assisted radical prostatectomy virtual reality simulation. Int J Med Robot. 2021;17(2):e2195.
62. Peng W, Xing Y, Liu R, Li J, Zhang Z. An automatic skill evaluation framework for robotic surgery training. International Journal of Medical Robotics and Computer Assisted Surgery. 2019;15(1):e1964.
63. Dubin AK, Julian D, Tanaka A, Mattingly P, Smith R. A model for predicting the GEARS score from virtual reality surgical simulator metrics. Surg Endosc. 2018;32(8):3576-81.
64. Zia A, Essa I. Automated surgical skill assessment in RMIS training. International Journal of Computer Assisted Radiology and Surgery. 2018;13(5):731-9.
65. Wang Z, Fey AM, editors. SATR-DL: Improving Surgical Skill Assessment And Task Recognition In Robot-Assisted Surgery With Deep Neural Networks2018: IEEE.
66. Watkinson W, Raison N, Abe T, Harrison P, Khan S, Van der Poel H, et al. Establishing objective benchmarks in robotic virtual reality simulation at the level of a competent surgeon using the RobotiX Mentor simulator. Postgrad Med J. 2018;94(1111):270-7.
67. Chen J, Oh PJ, Cheng N, Shah A, Montez J, Jarc A, et al. Use of Automated Performance Metrics to Measure Surgeon Performance during Robotic Vesicourethral Anastomosis and Methodical Development of a Training Tutorial. J Urol. 2018;200(4):895-902.
68. Hung AJ, Chen J, Che Z, Nilanon T, Jarc A, Titus M, et al. Utilizing Machine Learning and Automated Performance Metrics to Evaluate Robot-Assisted Radical Prostatectomy Performance and Predict Outcomes. J Endourol. 2018;32(5):438-44.
69. Fard MJ, Ameri S, Darin Ellis R, Chinnam RB, Pandya AK, Klein MD. Automated robot-assisted surgical skill evaluation: Predictive analytics approach. The International Journal of Medical Robotics and Computer Assisted Surgery. 2018;14(1):e1850.
70. 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. Surgical Endoscopy. 2018;32(1):62-72.
71. Raison N, Ahmed K, Fossati N, Buffi N, Mottrie A, Dasgupta P, et al. Competency based training in robotic surgery: benchmark scores for virtual reality robotic simulation. BJU International. 2017;119(5):804-11.
72. Vedula SS, Malpani A, Ahmidi N, Khudanpur S, Hager G, Chen CC. Task-Level vs. Segment-Level Quantitative Metrics for Surgical Skill Assessment. J Surg Educ. 2016;73(3):482-9.
73. Tanaka A, Graddy C, Simpson K, Perez M, Truong M, Smith R. Robotic surgery simulation validity and usability comparative analysis. Surg Endosc. 2016;30(9):3720-9.
74. Perrenot C, Perez M, Tran N, Jehl JP, Felblinger J, Bresler L, et al. The virtual reality simulator dV-Trainer(®) is a valid assessment tool for robotic surgical skills. Surg Endosc. 2012;26(9):2587-93.
75. Kumar R, Jog A, Malpani A, Vagvolgyi B, Yuh D, Nguyen H, et al. Assessing system operation skills in robotic surgery trainees. Int J Med Robot. 2012;8(1):118-24.
76. McDonough P, Peterson A, Brand T. Initial validation of the ProMIS surgical simulator as an OBJECTIVE measure of robotic task performance. Journal of Urology. 2010;183(4 SUPPL. 1):e515.

77. Judkins TN, Oleynikov D, Stergiou N. Objective evaluation of expert performance during human robotic surgical procedures. J Robot Surg. 2008;1(4):307-12.
78. Mark Knab L, Zenati MS, Khodakov A, Rice M, Al-Abbas A, Bartlett DL, et al. Evolution of a Novel Robotic Training Curriculum in a Complex General Surgical Oncology Fellowship. Ann Surg Oncol. 2018;25(12):3445-52.
79. Peng W, Xing Y, Liu R, Li J, Zhang Z. An automatic skill evaluation framework for robotic surgery training. Int J Med Robot. 2019;15(1):e1964.
80. Siddiqui NY, Galloway ML, Geller EJ, Green IC, Hur HC, Langston K, et al. Validity and reliability of the robotic Objective Structured Assessment of Technical Skills. Obstet Gynecol. 2014;123(6):1193-9.
81. Trinh L, Mingo S, Vanstrum EB, Sanford DI, Aastha, Ma R, et al. Survival Analysis Using Surgeon Skill Metrics and Patient Factors to Predict Urinary Continence Recovery After Robot-assisted Radical Prostatectomy. European Urology Focus. 2022;8(2):623-30.
82. Witthaus MW, Farooq S, Melnyk R, Campbell T, Saba P, Mathews E, et al. Incorporation and validation of clinically relevant performance metrics of simulation (CRPMS) into a novel full-immersion simulation platform for nerve-sparing robot-assisted radical prostatectomy (NS-RARP) utilizing three-dimensional printing and hydrogel casting technology. BJU International. 2020;125(2):322-32.
83. Frederick PJ, Szender JB, Hussein AA, Kesterson JP, Shelton JA, Anderson TL, et al. Surgical Competency for Robot-Assisted Hysterectomy: Development and Validation of a Robotic Hysterectomy Assessment Score (RHAS). Journal of Minimally Invasive Gynecology. 2017;24(1):55-61.
84. Cowan A, Chen J, Mingo S, Reddy SS, Ma R, Marshall S, et al. Virtual Reality vs Dry Laboratory Models: Comparing Automated Performance Metrics and Cognitive Workload during Robotic Simulation Training. Journal of Endourology. 2021;35(10):1571-6.
85. Nguyen JH, Chen J, Marshall SP, Ghodoussipour S, Chen A, Gill IS, et al. Using objective robotic automated performance metrics and task-evoked pupillary response to distinguish surgeon expertise. World J Urol. 2020;38(7):1599-605.
86. Vedula SS, Malpani A, Ahmidi N, Khudanpur S, Hager G, Chen CCG. Task-Level vs. Segment-Level Quantitative Metrics for Surgical Skill Assessment. Journal of Surgical Education. 2016;73(3):482-9.
87. Rice MK, Hodges JC, Bellon J, Borrebach J, Al Abbas AI, Hamad A, et al. Association of Mentorship and a Formal Robotic Proficiency Skills Curriculum With Subsequent Generations' Learning Curve and Safety for Robotic Pancreaticoduodenectomy. JAMA Surg. 2020;155(7):607-15.
88. Altok M, Achim MF, Matin SF, Pettaway CA, Chapin BF, Davis JW. A decade of robot-assisted radical prostatectomy training: Time-based metrics and qualitative grading for fellows and residents. Urol Oncol. 2018;36(1):13.e9-.e25.
89. Chowriappa AJ, Shi Y, Raza SJ, Ahmed K, Stegemann A, Wilding G, 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):561-9.
90. Møller SG, Dohrn N, Brisling SK, Larsen JCR, Klein M. Laparoscopic Versus Robotic-assisted Suturing Performance Among Novice Surgeons: A Blinded, Cross-Over Study. Surg Laparosc Endosc Percutan Tech. 2020;30(2):117-22.
91. Dilley J, Singh H, Pratt P, Omar I, Darzi A, Mayer E. Visual behaviour in robotic surgery-Demonstrating the validity of the simulated environment. International Journal of Medical Robotics and Computer Assisted Surgery. 2020;16(2):e2075.

92. 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. 2018;32(1):526-35.
93. Alrasheed T, Liu J, Hanasono MM, Butler CE, Selber JC. Robotic microsurgery: validating an assessment tool and plotting the learning curve. Plast Reconstr Surg. 2014;134(4):794-803.
94. Lyman WB, Passeri MJ, Murphy K, Siddiqui IA, Khan AS, Iannitti DA, et al. An objective approach to evaluate novice robotic surgeons using a combination of kinematics and stepwise cumulative sum (CUSUM) analyses. Surg Endosc. 2021;35(6):2765-72.
95. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273-8.
96. Mazzone E, Puliatti S, Amato M, Bunting B, Rocco B, Montorsi F, et al. A Systematic Review and Meta-analysis on the Impact of Proficiency-based Progression Simulation Training on Performance Outcomes. Ann Surg. 2021;274(2):281-9.
97. Lam K, Chen J, Wang Z, Iqbal FM, Darzi A, Lo B, et al. Machine learning for technical skill assessment in surgery: a systematic review. npj Digital Medicine. 2022;5(1):24.
98. Hung AJ, Chen J, Jarc A, Hatcher D, Djaladat H, Gill IS. Development and Validation of Objective Performance Metrics for Robot-Assisted Radical Prostatectomy: A Pilot Study. J Urol. 2018;199(1):296-304.
99. Bruce N, Pope D, Stanistreet D. Quantitative methods for health research: A practical interactive guide to epidemiology and statistics. West Sussex: Wiley; 2008 2008//.
100. https://www.rcsed.ac.uk/media/683822/rcsed-robotics-guidance-document-final.pdf.