

Review

Open Access



Systematic review on training models for partial nephrectomy

Rui J. Farinha^{1,2} , Elio Mazzone³, Marco Paciotti⁴, Alberto Breda⁵, James Porter⁶, Kris Maes², Ben Van Cleynenbreugel^{7,8}, Jozef Vander Sloten⁹, Alexandre Motttrie^{1,10}, Anthony G. Gallagher^{1,11,12}

¹Department of Research & Skills Development, Orsi Academy, Melle 9090, Belgium.

²Department of Urology, Hospital da Luz, Lisbon 1500-650, Portugal.

³Division of Oncology/Unit of Urology, Soldera Prostate Cancer Lab, URI, IRCCS Ospedale San Raffaele, Milan 20132, Italy.

⁴Department of Urology, Humanitas Clinical and Research Center IRCCS, Humanitas University, Rozzano, Milan 20132, Italy.

⁵Department of Urology, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona 08025, Spain.

⁶Swedish Urology Group, Swedish Medical Center, Seattle, WA 98107, USA.

⁷Department of Urology, University Hospitals Leuven, Leuven 3000, Belgium.

⁸Department of Development and Regeneration, KU Leuven, Leuven 3000, Belgium.

⁹Department of Mechanical Engineering, Section of Biomechanics, KU Leuven, Leuven 3000, Belgium.

¹⁰Department of Urology, Onze-Lieve-Vrouw Ziekenhuis, Aalst 9300, Belgium.

¹¹Faculty of Medicine, KU Leuven, Leuven 3000, Belgium.

¹²Faculty of Life and Health Sciences, Ulster University, Derry BT16DN, Northern Ireland, UK.

Correspondence to: Rui J. Farinha, Department of Research & Skills Development, Orsi Academy, Proefhoevestraat 12, Melle 9090, Belgium. E-mail: ruifarinhaurologia@gmail.com

How to cite this article: Farinha RJ, Mazzone E, Paciotti M, Breda A, Porter J, Maes K, Van Cleynenbreugel B, Vander Sloten J, Motttrie A, Gallagher AG. Systematic review on training models for partial nephrectomy. *Mini-invasive Surg* 2023;7:38. <https://dx.doi.org/10.20517/2574-1225.2023.50>

Received: 27 Apr 2023 **First Decision:** 28 Sep 2023 **Revised:** 14 Nov 2023 **Accepted:** 20 Nov 2023 **Published:** 27 Nov 2023

Academic Editors: Sey Kiat Terence Lim, Hendrik Van Poppel **Copy Editor:** Dong-Li Li **Production Editor:** Dong-Li Li

Abstract

Robot-assisted partial nephrectomy (PN) is a complex and index procedure with a difficult learning curve that urologists need to learn how to perform safely. We systematically evaluated the development and validation evidence underpinning PN training models (TMs) by extracting and reviewing data from PubMed, Cochrane Library Central, EMBASE, MEDLINE, and Scopus databases from inception to April 2023. The level of evidence was assessed using the Oxford Center for Evidence-Based Medicine. Of the 331 screened articles, 14 cohort studies were included in the analysis. No randomized controlled trials were found, and the heterogeneous nature of the models, study groups, task definitions, and subjectivity of the metrics used were transversal to all studies. All the models were rated good for realism and usefulness as training tools. Methodological discrepancies preclude definitive conclusions regarding the construct validation. No discriminative or predictive validation evidence was reported, nor were there comparisons between an experimental group trained with a TM and a control group. The



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



previous findings stand for the low level of evidence supporting the efficacy of the described TMs in the acquisition of skills required to safely perform PN.

Keywords: Surgical training, robot-assisted partial nephrectomy, construct validation, training model

INTRODUCTION

The difficult learning curve of laparoscopy^[1-3] and the advent of robotic surgery reinforced this transition and led to an exponential increase in the number of robot-assisted partial nephrectomy (RAPN) procedures performed. This is a complex and index procedure that urologists need to learn how to perform safely and has a difficult learning curve that requires a step-by-step training process. RAPN has several critical steps and requires the need to obtain negative surgical margins and control bleeding to avoid a potentially life-threatening hemorrhage^[4,5].

The introduction of surgical innovations and the need to ensure patient safety motivated international experts to develop structured training programs^[6,7] with validated curricula that include acquiring procedural skills in laboratory training models (TMs) and not simply relying on caseload. Rather, the goal necessitates demonstration of a proficiency benchmark in the skills laboratory before performing the procedure on a patient^[6].

Having access to a training center with animal-based *ex-* or *in-vivo* TMs might be the best option^[7]. Unfortunately, most trainees do not have access to this type of training facility, and since many hospitals cannot afford to purchase a robotic platform specifically for training purposes, 3D printed models and virtual reality (VR) simulators are considered cost-effective solutions for the acquisition of partial nephrectomy (PN) procedural skills.

Skills acquired using TMs can be transferred to the skill level required for safe surgical practice^[8], especially if surgeons are enrolled in a proficiency-based progression (PBP) training program for PN^[9]. However, this approach is contingent on high-level validation evidence supporting the use of a TM^[10].

This review sought to evaluate the type and level of validation in the literature on the efficacy of existing PN TMs and demonstrate the skill acquisition and performance levels required for safe surgical practice.

MATERIALS AND METHODS

Search strategy

A systematic review of the literature was conducted using the PubMed, Cochrane Library Central, EMBASE, MEDLINE, and Scopus databases. We searched from the inception of the databases until April 2023. All references in the included papers on TMs were also screened. The keywords used for this research were “Partial nephrectomy AND Training models”. The scope of this research was limited to the English language. This systematic review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocol (PRISMA-P) guidelines^[11,12].

Data extraction and analysis

After identifying all eligible studies, two independent reviewers (Farinha RJ and Mazzone E) screened all titles and abstracts or full texts for further clarification and inclusion. Literature reviews, editorial commentaries, and non-PN TM studies were excluded from the initial screening. Randomized controlled trials (RCTs) and nonrandomized observational studies (cohort studies) on validity and skill transfer from

the TM to clinical PN were included. Other inclusion criteria were the use of objective metrics to measure task execution or subjective assessments of PN performance using the scores of global evaluative assessment of robotic skills (GEARS) or global operative assessment of operative skills (GOALS)^[9,13-29].

Disagreements regarding eligibility were resolved by discussion between the two investigators until a consensus was reached regarding the studies to be included. The level of evidence was assigned according to the Oxford Center for Evidence-based Medicine definitions^[30]. This article does not contain any studies involving animals performed by any of the authors.

RESULTS

Study selection

Figure 1 shows the flow of studies through the screening process. A total of 331 papers were blindly screened by two reviewers (Farinha RJ and Mazzone E) by reading all titles and abstracts, with 16 of these records included for further evaluation based on predefined eligibility criteria. At this point, the final evaluation for inclusion in the quantitative analysis was carried out by three reviewers (Gallagher AG, Farinha RJ, and Mazzone E), who selected 14 manuscripts.

Evidence synthesis

Training models

The final screened manuscripts included four animal-based, eight 3D printed, and two VR TM studies for PN procedural training. Animal TMs were used *in vivo*^[14], but more commonly, *ex vivo*^[9,15,16] models employing porcine kidneys were employed. Pseudo-tumors were created either through percutaneous injection of liquid plastic^[14], gluing a styrofoam ball to the renal parenchyma^[30], or simply demarcating an area to be resected^[9,15]. The pseudo-tumoral areas were established in accessible portions of the renal parenchyma, with sizes varying between 2 and 3.8 cm^[9,14,16,31], and perfusion was emulated in two of the models^[16,32] [Table 1].

The 3D printed models were based on computed tomography (CT) or magnetic resonance imaging (MRI) images of real patients and, therefore, were patient-specific. Usually, a mold is 3D printed^[17,19,23,25] and filled with polyvinyl acetate (PVA-C)^[18,19], silicone^[17,23,25,26], agarose gel^[24], or N-composite gel^[29]. Being used for preoperative rehearsal^[17,19,23-26], they included pseudo-tumors with 1.5 to 4.7 cm, vascular structures for “blood” perfusion^[17,19] and sometimes other anatomical structures (i.e., renal hilum, pelvicalyceal system, colon, spleen, and anterior abdominal wall^[19,26,29] [Table 1].

VR and augmented reality (AR) technologies were used to develop PN simulation platforms^[27,28], with the goal of teaching surgical anatomy (knowledge), technical skills, and operative steps (basic and procedural skills). Using the CT images of patients, preoperative rehearsal was possible^[28], and the integration of computer-based performance metrics allowed the assessment of surgical performance^[27] [Table 1].

The most common emulated core tasks were tumor excision^[9,14,16-19,23,24,26,29,31,32] and renorrhaphy^[15,17,19,23,24,27,29,32]. The 3D TMs also emulated the control of hemostasis^[14], renal hilum dissection^[19], renal artery clamping^[17,19], instrument choice^[17], colon mobilization^[19], port placement^[17], intraoperative ultrasound^[17,19], and specimen entrapment^[17].

Studies

The level of evidence of all included studies was $\geq 3b$; different face, content, and construct validation studies were identified, and a summary is presented in Table 2.

Table 1. Partial nephrectomy training models

Studies	Model	Surgery	Material	Tumor size (cm)	Extra features
Hidalgo <i>et al.</i> ^[16]	Animal	LPN	Liquid plastic	2	Perfusion
Yang <i>et al.</i> ^[32]	Animal	LPN	Demarcation area	2	Perfusion
Hung <i>et al.</i> ^[18]	Animal	RAPN	Styrofoam	3.8	No
Chow <i>et al.</i> ^[11]	Animal	RAPN	Demarcation area	2.5	No
Fernandez <i>et al.</i> ^[20]	3D	LPN	PVA-C	1.5	No
Golab <i>et al.</i> ^[25]	3D	LPN	Silicone	n/a	No
Monda <i>et al.</i> ^[19]	3D	RAPN	Silicone	4	Surgical tubing to emulate renal hilum
Ghazi <i>et al.</i> ^[21]	3D	RAPN	PVA-C	4.2	Hilar hollow structures; pelvicalyceal system; retroperitoneal structures; colon; spleen; anterior abdominal wall
Maddox <i>et al.</i> ^[26]	3D	RAPN	Agarose gel	4.7	No
Von Rundstedt <i>et al.</i> ^[27]	3D	RAPN	Silicone	4	No
Glybochko <i>et al.</i> ^[28]	3D	LPN	Silicone	n/a	Vascular system; pelvicalyceal system
Ohtake <i>et al.</i> ^[33]	3D	LPN	N-composite gel	n/a	Pelvicalyceal system
Makiyama <i>et al.</i> ^[30]	VR	LPN	Software	n/a	Surrounding structures
Hung <i>et al.</i> ^[29]	VR	RAPN	Software	n/a	Surrounding structures

LPN: Laparoscopic partial nephrectomy; PVA-C: polyvinyl acetate; RAPN: robot-assisted partial nephrectomy; VR: virtual reality.

Table 2. Validation studies

Studies	Face	Content	Construct	Concurrent	Feasibility	Predictive	Transfer of skills
Hidalgo <i>et al.</i> ^[16]	Yes	Yes	No	No	No	No	No
Yang <i>et al.</i> ^[32]	Yes	Yes	No	No	No	No	No
Hung <i>et al.</i> ^[44]	Yes	Yes	Yes	No	No	No	No
Chow <i>et al.</i> ^[11]	Yes	Yes	Yes	No	No	No	No
Fernandez <i>et al.</i> ^[20]	Yes	Yes	No	No	No	No	No
Golab <i>et al.</i> ^[25]	No	No	No	No	No	No	No
Monda <i>et al.</i> ^[19]	Yes	Yes	Yes	No	No	No	No
Ghazi <i>et al.</i> ^[21]	Yes	Yes	Yes	No	No	No	No
Maddox <i>et al.</i> ^[26]	No	No	No	No	Yes	No	No
Rundstedt <i>et al.</i> ^[27]	No	No	No	No	Yes	No	No
Glybochko <i>et al.</i> ^[28]	Yes	No	No	No	Yes	No	No
Ohtake <i>et al.</i> ^[33]	Yes	Yes	Yes	No	No	No	No
Makiyama <i>et al.</i> ^[30]	Yes	Yes	No	No	Yes	No	No
Hung <i>et al.</i> ^[29]	Yes	Yes	Yes	Yes	No	No	No

Face validity

Experts assess face validity by determining whether a test measures what it is intended to^[18]. When applied to surgical simulators, this is equivalent to realism. Four animals^[9,14,16,31], five 3D printed^[17-19,26,29], and two VR-TM studies reported face validity results^[27,28]. In three animal^[9,16,31], four 3D^[17-19,29], and two VR^[27,28] TM studies, face validity was evaluated by all participants, including novices, without any surgical experience and was rated exclusively by experts in one animal^[14] and one 3D TM study^[26]. The participants answered a questionnaire immediately^[9], several days^[31], and one week after their use^[14]. One animal study used four

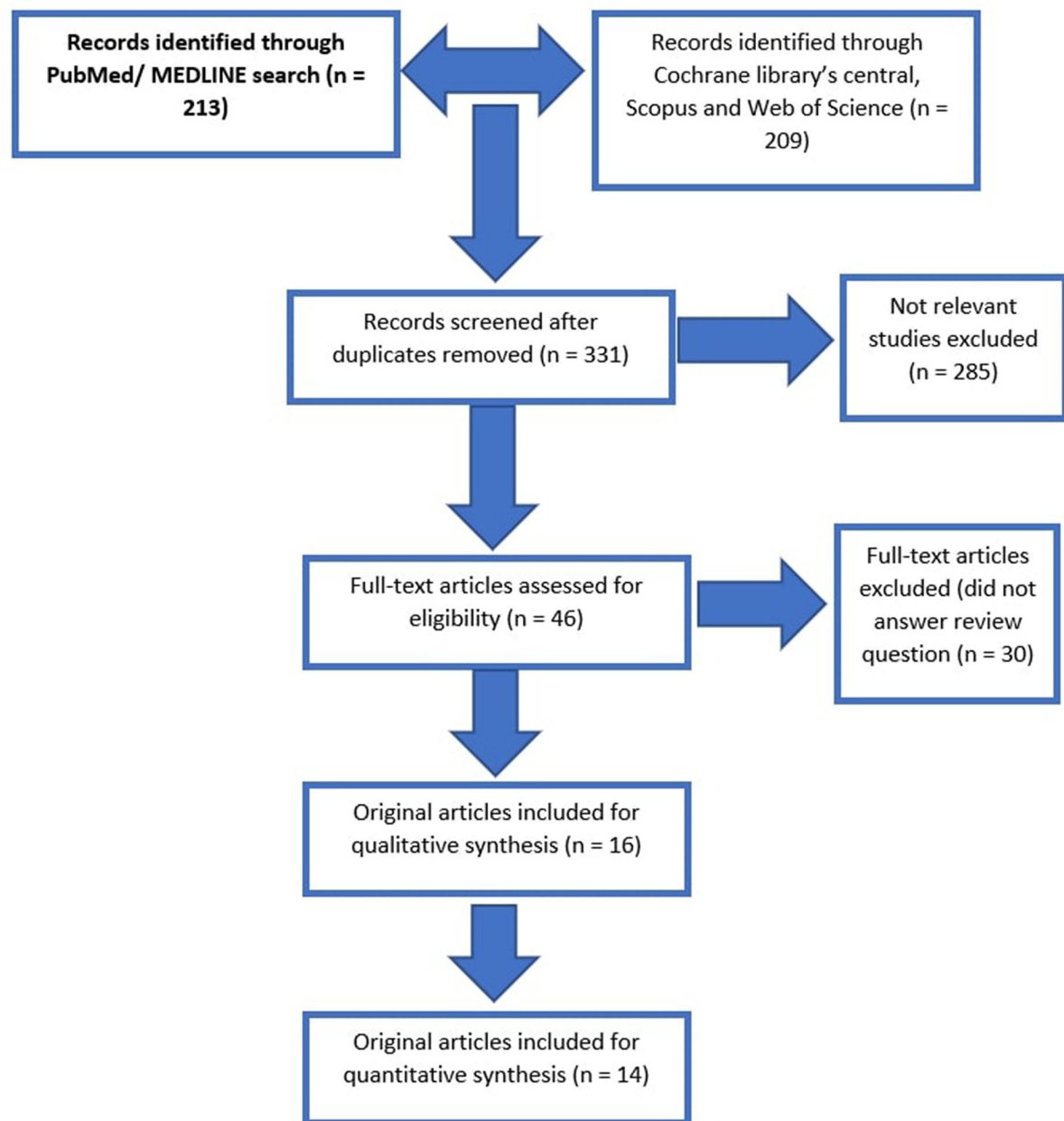


Figure 1. Study selection process, according to the PRISMA Statement. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

questions that were assessed using a ten-point Likert scale, where 96% of the participants reported an enhancement and no hindering of their learning experience^[14]. By answering one question on a ten-point Likert scale, one animal study reported that all participants considered the model helpful in improving their confidence and skills in performing PN^[31]. In another study, the experts rated the TM as “very realistic” [median score 7/10, range (6-9)]^[27], and in another study, the model was rated as having contributed to their skill (4/5) and confidence (4.1 out of 5) in performing robotic surgery^[9].

A questionnaire was completed immediately after using the 3D TMs^[17-19,29] with a five-^[18,29] or 100-point Likert scale^[17], and two studies did not report the type of assessment scale used^[19,26]. All models were reported as having “good realism”^[18,26] concerning the form and structure of the kidney and as being “high”^[29] or even superior to porcine or cadaveric models^[19]. One study reported detailed face validity data for the model’s overall feel (mean 79.2), usefulness (mean 90.7), realism for needle driving (mean 78.3), cutting (mean 78.0), and visual representation (mean 78.0)^[17].

The VR TMs were evaluated with a questionnaire immediately after the model’s use^[27,28], and the questions were scored using a five-^[27] or ten-point Likert scale^[28]. One VR TM study reported that the full-length AR platform was very realistic (median 8/10, range 5-10) compared to the *in vivo* porcine model (median 9/10, range 7-10, $P = 0.07$)^[27], and another study reported a mean score for anatomical integrity of 3.4 (± 1.1) using a five-point analog visual scale^[28].

Content validity

Content validity measures whether skills training on a simulator is appropriate and correct, classifying the model’s usefulness as a training tool^[18]. Our research identified content validity studies on animals^[9,14-16], four on 3D printed^[17-19,29] and two on VR TMs^[27,28]. All participants, including novices (without any surgical experience), were evaluated in three animal^[9,16,31], four 3D^[17-19,29], and two VR^[27,28] TM studies and were exclusively assessed by experts in one animal study^[14].

In animal TM studies, qualitative evaluations are derived from unspecified questionnaires. Either participants found the model “helpful”^[31], rated it as an “extremely useful” training tool for residents (9/10; range 7.5-10) and Fellows (9/10; range 7-10), although less so for experienced robotic surgeons (5/10; range 3-10), or high ratings of usefulness (4/5) were attributed by participating residents^[9]. In one study, TM was evaluated exclusively by experts who considered it to enhance their learning experience (96%)^[14].

In 3D TM studies, unspecified questionnaires use qualitative evaluation and Likert scales to assess and report results on content validity. One model is “recommended as a teaching tool” for residents and fellows^[18]. Another was considered “useful as a training tool” by 93.7% of the participants^[19], and another study reported a total content score of 4.2 using a five-point Likert scale^[29].

Using a non-validated questionnaire and a 0-100 Likert scale anchored to useless-useful, one model reached 90.7 for overall usefulness for training, being considered most useful “for trainees to obtain new technical skills” (mean score 93.8) and less useful “for trainees to improve existing technical skills” (mean score 85.7)^[17]. The only study in this group of TMs, in which the assessment was exclusively performed by experts, did not report data on content validity^[26].

Using an unspecified questionnaire, experts rated the procedure-specific VR renorrhaphy exercise as highly useful for training residents and fellows, although less useful for experienced robotic surgeons new to RAPN. The model was highly rated for teaching surgical anatomy (median 9/10, range 4-10) and procedural steps (8.5/10, range 4-10). Technical skills training was rated slightly lower, although still favorably (7.5/10, range 1 to 10)^[27]. Using a visual analog scale (score range 1-5), the surgeons evaluated the utility of the simulations, attributing a score of 4.2 (± 1.1)^[28].

Construct validity

Construct validity denotes the ability of a simulator to differentiate between experts and novices on given tasks^[18], thereby providing clinically meaningful assessments^[18]. Our review identified six cohort studies on construct validity^[9,16,17,19,27,29].

Fifty-eight participants were enrolled in two animals^[9,16], 83 participants in three 3D^[17,19,29], and 42 in one VR/AR TM study^[27].

The study participants were medical students, residents, fellows, and attending surgeons. The criteria used to classify them into “novice”, “intermediate”, and “expert” groups varied between studies^[16,17,19,27,29]. For example, the definition of “expert”, as a surgeon with > 100^[16,27] or > 150 console cases^[19], was based on the number of surgical cases completed^[16,19,27,29]. The experiences of the different enrolled cohorts varied considerably, including subjects without any surgical experience^[17,27]. Comparisons between two groups with a clear discrimination of surgical experience (novices and experts)^[29], and three groups without a clear difference in experience (novices, intermediates, and experts)^[9,16,17,19] were identified.

Photo or video recordings of the surgeon’s performance were collected, and experts were blinded to the experience level and the surgeon performing the task. The metrics used varied from GEARS^[9,19,27], GOALS^[16,29], and clinically relevant outcome measures (CROMS)^[19] to different operation-specific metrics, namely, time (renal artery clamping^[17,19], tumor excision^[9,34], total operative^[9,16], and console time^[19]), estimated blood loss^[19], preserved renal parenchyma^[17], surgical margin status^[16,17,19,29], maximum gap between the two sides of the incision^[29], total split length^[29], and quality of PN (scored on a Likert scale)^[9]. In one animal model, instrument and camera awareness and the precision of instrument action were subjectively scored using a Likert scale^[27]. Built-in algorithm software metrics were used in one VR TM, scoring instrument collisions, instrument time out of view, excessive instrument force, economy of motion, time to task completion, and incorrect answers^[27] [Table 3].

Concurrent validity

One AR/VR simulator study compared the performance of experts on a virtual and an *in vivo* porcine renorrhaphy task. It was found to have equal realism and high usefulness for teaching anatomy, procedural steps, and training technical skills of residents and fellows, although less so for experienced robotic surgeons new to RAPN^[27].

Kane’s framework

Following Kane’s framework^[18] of the validation process, focusing on decisions and consequences, the fragilities of the analyzed studies become more obvious. The proposed use of different models varies from developing and testing them to evaluating distinct levels of validation^[9,14,16-29]. The type of scoring used is based on the timing of various steps of the emulated procedure and/or using Likert scales, such as GEARS or GOALS^[9,13,14,16-29,31].

None of the studies generalized the test results to other tasks. Several authors report their models as realistic and useful training tools for residents and fellows, although they are usually not considered highly beneficial for training consultants^[9,14,16-29,31]. The implications of using diverse models differ across studies. Generally considered an effective surgical education/training tool to learn key steps of PN and develop advanced laparoscopic/robotic skills, they are associated with fewer logistic concerns. This is due to their lack of necessity for dedicated teaching robots or wet/laboratory facilities [Table 4].

DISCUSSION

The aviation industry established the safety benefit of training on simulators many decades ago^[35], inspiring surgeons to pursue their training in the laboratory before entering the operating room^[36,37]. Skills acquired

Table 3. Construct validation studies

Authors	Participants enrolled			Data used		Assessor	Scales
	Novices	Intermediates	Experts	Photos	Videos		
Hung et al. ^[18]	24 (O CC)	9 (< 100 CC)	13 (> 100 CC)	Yes		Two experts	Likert scale
Chow et al. ^[11]	6 (PGY 2-3)		6 (PGY 4-5)		Yes	Three experts	GOALS; I/C A; PIA
Monda et al. ^[19]	12 4 MS + 8 (2nd/3rd) YR	6 4th and 5th YR	6 (3 fel. + 3 cons.)		Yes	5 FTFM	GEARS
Ghazi et al. ^[21]	27 (22 res. + 5 fel.; < 30 TRC)		16 (cons; > 150 UTRC)		Yes	2 FTAS (> 200 RAPN)	GEARS
Ohtake et al. ^[33]	8 (< 20 LP)		8 (> 20 LP)		Yes		GEARS/CROMS
Hung et al. ^[29]	15 (no ST)	13 (< 100 CC)	14 (> 100 CC)		Yes	One expert	GOALS/TPT
	92	28	63		Yes		

CC: Console cases; cons.: consultant; CROMS: clinically relevant outcome measures; fel.: fellows; FTAS: fellowship trained attending surgeons; FTFM: fellowship trained faculty members; GEARS: global evaluative assessment of robotic skills; GOALS: global operative assessment of operative skills; I/C A: instrument/camera awareness; LP: laparoscopic procedure; MS: medical students; PGY: post graduate year; PIA: precision of instrument action; RAPN: robot-assisted partial nephrectomy; res.: residents; ST: surgical training; TPT: total procedure time; UTRC: upper tract robotic cases; YR: year resident.

using TMs can be transferred to the performance level required for safe surgical practice^[8], especially if surgeons are enrolled in a PBP training program for PN^[10], although this recommendation is contingent on a high level of evidence^[10].

As a reference procedure that urologists need to learn with a difficult learning curve and potentially life-threatening complications, the acquisition of skills for the performance of a safe PN should start in the skills laboratory. This review aimed to evaluate the type and level of validation evidence for the efficacy of existing PN TMs in acquiring and transferring surgical skills to the performance level required for safe surgical performance. No RCTs were found among the reviewed studies. Fourteen cohort studies on PN TMs based on animal tissue, 3D printing, and VR/AR technology were identified. Using the classification developed by the Oxford Center for Evidence-Based Medicine, the level of evidence assessed was low^[30].

Training models

Animal TMs closely emulate human tissues, allowing trainees to understand anatomical structures, natural tissue consistency, and movement during dissection and suturing. These are critical features for training in tumor excision and renorrhaphy. The reviewed studies used different substances to create pseudo-tumors of a consistent size. Although no cost-effective studies have been conducted, these models were found to be economical and widely available.

Several potential advantages were identified with 3D printed TMs. They were derived from the patient's CT or MRI images and were, therefore, patient-specific. Furthermore, they provide the potential benefits of preoperative rehearsal. The technology used to print the mold produced durable, reliable, and repeatable models, and the created phantoms accurately represented the patient's anatomy and diverse tumor geometries.

Different substances were used to fill the mold to produce the final model. Silicone represented the kidney tissue in terms of tear strength, but PVA-C was the most frequently used^[17,23,25,26]. The latter closely resembled real tissue, allowing the addition of enhancing agents (gadolinium and barium), providing effective imaging by CT or MRI, which could be recycled.

Table 4. Kane framework

	Proposed use (decision)	(Type of) Scoring	Generalization	Extrapolation	Implications
Hidalgo <i>et al.</i> ^[16]	Develop and test an <i>in-vivo</i> porcine LPN TM to teach LPN	Use time as a metric in different steps	None identified	The model enhances the learning experience	Participants endorsed application of the model as an effective surgical educational tool
Yang <i>et al.</i> ^[32]	Develop and test an <i>ex-vivo</i> porcine LPN TM to teach LPN	Used operation-specific and time metrics Measured learning curve and quality of PN	None identified	Trainees found the model helpful, increased confidence, and improved skills in LPN	Authors consider the model useful for learning key steps of PN and developing advanced laparoscopic suture-repairing skills
Hung <i>et al.</i> ^[18]	Evaluate face, content, and construct validities of <i>ex-vivo</i> RAPN TM	Used questionnaires to assess realism and utility as training tools Video recordings were assessed by three experts Use time, operation-specific metrics, and GOALS	None identified	Experts rated the model high in realism and as a training tool for residents and fellows. Limited training role for expert surgeons	A model appropriate for resident and fellow training
Chow <i>et al.</i> ^[11]	Assess validity and effectiveness of an <i>ex-vivo</i> porcine TM	Used time and GEARS Video recorded performances Blinded assessors	None identified	Improved skills, shortened the learning curve, and increased operator confidence	Use of this model in a urology residency curriculum
Fernandez <i>et al.</i> ^[20]	Evaluate the materials model for PN kidney tumors	Likert-scale to rate quality and realism of renal tumor model Evaluated operation-specific and time metrics Evaluated learning curve measuring time	None identified	Rated as "good" realism Participants considered the model helpful in learning to perform LPN Good teaching tool for residents and fellows to learn technical skills of the LPN	PVA-C use was less expensive and entailed fewer logistic concerns than those associated with the animal model
Golab <i>et al.</i> ^[25]	Create individual silicone models for training LPN	Used time as metrics	None identified	Improved actual surgery Reduced the need for/duration of intraoperative renal ischemia	Producing these models brings new possibilities for laparoscopic education
Monda <i>et al.</i> ^[19]	Assess face, content, and construct validity of a RAPN training model	Evaluated usefulness and realism of the model as a training tool, Performance measured using operation-specific metrics, NASA-TLX and GEARS Video performance recorded and blinded assessments by experts	None identified	Experts gave high ratings for realism and usefulness Differentiated surgical performance of groups' expertise Evidenced learning curve	Novel and economic methods of manufacturing silicone models Useful for trainees to gain fundamental surgical skills in RALPN
Ghazi <i>et al.</i> ^[21]	Simulation platform for RAPN	Used CROMS and GOALS Evaluated realism ratings and training effectiveness	None identified	Rated by experts as superior to porcine or cadaveric models for replication of procedural steps Excellent at discriminating experts from novice performance	The model might lead to widespread use of procedural, patient-specific, individualized practice No need for dedicated teaching robots and wet-laboratory facilities
Maddox <i>et al.</i> ^[26]	Develop patient-specific kidney models for the purpose of pre-surgical resection and incorporation into simulation labs	No scoring. Compare clinical results between patients from the study and similar studies from a RAPN database	None identified	Patients who underwent the preoperative surgical model experienced lower estimated blood loss at the time of resection	Use of this type of model may decrease the slope of the learning curve and improve patient outcomes
von Rundstedt <i>et al.</i> ^[27]	Develop patient-specific pre-surgical simulation protocol for RALPN	Compare resection times between the model and the actual tumor in a patient-specific manner	None identified	Improved resection times Similar morphology and tumor volumes when compared with the real tumor Predict feasibility of RALPN within an acceptable ischemia time	Can assist in surgical decision-making, provide preoperative rehearsals, and improve surgical training

Glybochko <i>et al.</i> ^[28]	Evaluate effectiveness of personalized 3D printed models for pre-surgical planning	Used time-based metrics and blood loss	None identified	Elasticity and density similar to real kidney	Can contribute to improvement of surgical skills and facilitate selection of optimal surgical tactics
Ohtake <i>et al.</i> ^[33]	Examine effectiveness of the model as a tool for practicing LPN	Used Lickert-scale questionnaires to evaluate realism and utility as training tools Used GOALS to score performance Used procedure-specific metrics	None identified	Significant differences between novice and expert performance Improvement in the learning curve	Can be used daily as a training tool for LPN
Makiyama <i>et al.</i> ^[30]	Describe and validate a patient-specific simulator for laparoscopic surgery	Visual analog scales to assess anatomical integrity and utility and intraoperative confidence during subsequent surgical procedures	None identified	Reproduced patient anatomy High scores in the utility of simulations and surgeons' intraoperative confidence	Useful as a preoperative training tool Improvements still needed
Hung <i>et al.</i> ^[29]	Evaluate face, content, construct, and concurrent validity	Questionnaires to evaluate realism and usefulness for training Used GEARS and computer-based performance metrics	None identified	Differentiated performance of experts from non-experts Highly useful in training residents and fellows but less so for experienced surgeons Inferior utility in training compared with porcine Scored high to teach surgical anatomy and procedure steps	Although validated, several areas need improvement, particularly with the teaching of advanced technical skills

CROMS: Clinically relevant outcome measures; GEARS: global evaluative assessment of robotic skills; GOALS: global operative assessment of operative skills; LPN: laparoscopic partial nephrectomy; NASA-TLX: NASA-task load index; PN: partial nephrectomy; RALPN: robot assisted partial nephrectomy; RAPN: robot-assisted partial nephrectomy; TM: training model.

Although the preparation and use of 3D printed models were labor intensive, and monofilament sutures were recommended (e.g., braided sutures easily torn this material)^[18,19], they involved fewer logistic concerns than the use of animal models^[18,19]. They are simple, easy to set up, and likely have a practically indefinite shelf life. The price was reported in some studies, purporting its economic value, but the cost of the 3D printer was not considered^[17,19,23,26].

The feasibility of incorporation into a training course was the focus when selecting clinically relevant steps to emulate. Therefore, most of the 3D printed models focused on simulating tumor resection and renorrhaphy. Some models include other anatomical structures, potentially increasing their realism and educational value^[19,26,29].

The exponential increase in computing power over the last decade makes VR/AR TMs very promising. By including different teaching tasks, patient-specific TMs allow preoperative rehearsal. However, signal processing delays induce a lack of realistic tissue responsiveness during the dissection of tissue planes, tissue excision, suturing, knot tying, and bleeding, which significantly compromises the capacity of VR simulation to accurately emulate the PN procedure and thus their value as a training tool^[27,28].

Despite the advantages outlined herein, these TMs have several drawbacks. The need to optimize perfusion flow pressures, lack of hilar dissection, clamping, and hemostasis management were identified as potentially needing improvements. Overcoming these shortcomings will accelerate the evolution from basic benchtop and part-task trainers to the development of realistic and accurate recreation of an entire PN procedure, which would underpin effective surgical training.

Studies

The clinical differentiation of the study population was heterogeneous, and the skill level criteria used to differentiate novices, intermediates, and experts varied considerably between studies. These criteria were unclear, and expertise was defined based on the number of surgeries performed rather than the number of PNs performed by the surgeon.

The face and content validity studies used qualitative (i.e., based on Likert scales) questionnaires that did not appear to be supported by validation evidence^[9,19,29]. Responses were elicited from the participants in variable time frames, that is, up to one week after the use of the TM^[14]. Reports of high rates of realism and usefulness of training tool results were mainly obtained from experts' evaluations. Furthermore, some studies enrolled novice surgeons with slim-to-no PN operative experience^[9,18,29,31].

One study used photographs of the models and the tasks performed to complete the evaluation^[16]. The majority of the construct validity studies assessed video recordings^[9,17,19,27,29]. They used expert assessors who were blinded to the experience level and surgeon performing the task. Time was employed as the main metric despite evidence demonstrating that it has a weak association with performance quality^[38]. Only one concurrent validity study was conducted with one VR simulator, and no studies assessing the predictive aspect or transfer of skills were identified.

In the studies reviewed, Likert-type scales, such as GEARS and GOALS, were used to evaluate users' performance in the TMs, although it was consistently demonstrated that they produce unreliable assessment measures^[9,16,19,27,29,39]. No procedure-specific binary metrics were reported, and none of the tasks used performance errors as units of performance assessment. Furthermore, the methodology employed to train assessors in using the assessment scales was not reported, nor was an interrater reliability level.

All identified validation studies followed the nomenclature and methodology described by Messick^[40] and Cronbach^[41] rather than the framework described by Kane^[18], reporting data on face, content, construct, and concurrent validation instead of using Kane's validation processes (i.e., scoring, generalization, extrapolation, and implication)^[18]. In the "Scoring inference", the developed skill stations included different performance steps of the PN, and fairness was partially guaranteed by the production of standardized TMs. However, the main problem was that scoring predominantly used global rating scales with no reported attempts to demonstrate or deal with the issue of performance score reliability.

Furthermore, no effort was expended in the "Generalization inference" area. The items used to assess performance were ill-defined. The researchers did not evaluate the reproducibility of scores, nor did they investigate the magnitude of performance error; therefore, there was no identification of the sources of error.

The studies reviewed here investigated whether the test domains reflected key aspects of the real PN, but no analysis was performed to evaluate the relationship between the performance and real-world performance. The same can be said about the "Implications inference" theme. Although a weak evaluation of the impact

of the model's use on users was shown, no impact evaluation of its use was addressed outside the study population. Furthermore, no comparison between groups of users and non-users of TMs was undertaken, nor an analysis of relevant clinical outcomes was performed. All these observations make it very difficult to gather evidence supporting the decision to integrate these TMs into PN training programs.

Several fundamental flaws pervaded the reviewed studies. There was considerable heterogeneity in the materials used to build the TMs, a lack of comparisons between the different models, and objective binary metrics demonstrating skill improvement. Although cost was described in some studies, no cost-effectiveness data were reported, and the level of evidence to support their use for training purposes was weak. All these reasons preclude a recommendation for the adoption of these TMs in PN training programs.

Since TMs are a tool for delivering a metric-based training curriculum, future research should focus on the improvement of the models, and the starting point should be the development of objective, transparent, and fair procedural-specific metrics^[42]. A clear definition of expertise criteria, considering the performance level of the surgeons and not the number of surgeries performed, should be a main concern. Kane's framework for study validation should be used, and comparisons should be made between models and between study groups trained with and without the different TMs. Improvements will only emerge from the conjoined efforts of surgeons, human factor engineers, training experts, and behavioral scientists^[43].

CONCLUSION

This review substantiates the absence of well-designed validation studies on PN TMs and their inherently low level of scientific evidence. No RCTs or impact inferences were found to support the adoption of TMs in PN training curricula.

APPENDIX

Face validity: opinions, including of non-experts, regarding the realism of the simulator.

Content validity: opinions of experts about the simulator and its appropriateness for training.

Construct validity: (A) one group: ability of the simulator to assess and differentiate between the level of experience of an individual or group measured over time; (B) between groups: ability of the simulator to distinguish between different levels of experience.

Concurrent validity: comparison of the new model against the older and gold standard.

Predictive validity: correlation of performance with operating room performance.

DECLARATIONS

Authors' contributions

Study concept and design, analysis and interpretation, drafting of the manuscript, statistical analysis, administrative, technical or material support: Farinha RJ, Gallagher AG

Acquisition of data: Farinha RJ, Mazzone E, Paciotti M

Critical revision of the manuscript for important intellectual content: Farinha RJ, Breda A, Porter J, Maes K, Van Cleynenbreugel B, Vander Sloten J, Mottrie A, Gallagher AG

Supervision: Gallagher AG

Farinha RJ had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

All authors participated in the study, writing, and approval of the manuscript for submission and accept accountability, adhering to the International Committee of Medical Journal Editors requirements.

Availability of data and materials

All data were obtained from the published articles.

Financial support and sponsorship

The present research project has been conducted by Rui Farinha as part of his PhD studies in KU Leuven, Belgium, and of the ongoing project for the ERUS and ORSI Academy. For the design, research, data collection, analysis, and preparation of the manuscript, the funding was the following: none.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2023.

REFERENCES

1. Link RE, Bhayani SB, Allaf ME, et al. Exploring the learning curve, pathological outcomes and perioperative morbidity of laparoscopic partial nephrectomy performed for renal mass. *J Urol* 2005;173:1690-4. [DOI](#)
2. Gill IS, Kamoi K, Aron M, Desai MM. 800 Laparoscopic partial nephrectomies: a single surgeon series. *J Urol* 2010;183:34-41. [DOI](#)
3. Hanzly M, Frederick A, Creighton T, et al. Learning curves for robot-assisted and laparoscopic partial nephrectomy. *J Endourol* 2015;29:297-303. [DOI](#)
4. Patel HD, Mullins JK, Pierorazio PM, et al. Trends in renal surgery: robotic technology is associated with increased use of partial nephrectomy. *J Urol* 2013;189:1229-35. [DOI](#)
5. Alameddine M, Koru-Sengul T, Moore KJ, et al. Trends in utilization of robotic and open partial nephrectomy for management of cT1 renal masses. *Eur Urol Focus* 2019;5:482-7. [DOI](#) [PubMed](#)
6. Smith R, Patel V, Satava R. Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. *Int J Med Robot* 2014;10:379-84. [DOI](#) [PubMed](#)
7. 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;81:767-74. [DOI](#)
8. Ahmed K, Khan R, Motttrie A, et al. Development of a standardised training curriculum for robotic surgery: a consensus statement from an international multidisciplinary group of experts. *BJU Int* 2015;116:93-101. [DOI](#)
9. Raison N, Gavazzi A, Abe T, Ahmed K, Dasgupta P. Virtually competent: a comparative analysis of virtual reality and dry-lab robotic simulation training. *J Endourol* 2020;34:379-84. [DOI](#) [PubMed](#)
10. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 2002;236:458-64. [DOI](#) [PubMed](#) [PMC](#)
11. Chow AK, Wong R, Monda S, 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;35:878-84. [DOI](#) [PubMed](#)
12. Dawe SR, Windsor JA, Broeders JA, Cregan PC, Hewett PJ, Maddern GJ. A systematic review of surgical skills transfer after simulation-based training: laparoscopic cholecystectomy and endoscopy. *Ann Surg* 2014;259:236-48. [DOI](#) [PubMed](#)
13. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700. [DOI](#) [PubMed](#) [PMC](#)
14. 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;187:247-52. [DOI](#) [PubMed](#)
15. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 2005;190:107-13. [DOI](#)
16. Hidalgo J, Belani J, Maxwell K, et al. Development of exophytic tumor model for laparoscopic partial nephrectomy: technique and initial experience. *Urology* 2005;65:872-6. [DOI](#)
17. Yang B, Zhang ZS, Xiao L, Wang LH, Xu CL, Sun YH. A novel training model for retroperitoneal laparoscopic dismembered pyeloplasty. *J Endourol* 2010;24:1345-9. [DOI](#)

18. Hung AJ, Ng CK, Patil MB, et al. Validation of a novel robotic-assisted partial nephrectomy surgical training model. *BJU Int* 2012;110:870-4. DOI PubMed
19. Monda SM, Weese JR, Anderson BG, et al. Development and validity of a silicone renal tumor model for robotic partial nephrectomy training. *Urology* 2018;114:114-20. DOI PubMed PMC
20. Fernandez A, Chen E, Moore J, et al. First prize: a phantom model as a teaching modality for laparoscopic partial nephrectomy. *J Endourol* 2012;26:1-5. DOI PubMed
21. Ghazi A, Melnyk R, Hung AJ, 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;127:645-53. DOI PubMed
22. Hongo F, Fujihara A, Inoue Y, Yamada Y, Ukimura O. Three-dimensional-printed soft kidney model for surgical simulation of robot-assisted partial nephrectomy: a proof-of-concept study. *Int J Urol* 2021;28:870-1. DOI PubMed
23. Vitagliano G, Mey L, Rico L, Birkner S, Ringa M, Biancucci M. Construction of a 3D surgical model for minimally invasive partial nephrectomy: the urotrainer VK-1. *Curr Urol Rep* 2021;22:48. DOI PubMed
24. Melnyk R, Ezzat B, Belfast E, et al. Mechanical and functional validation of a perfused, robot-assisted partial nephrectomy simulation platform using a combination of 3D printing and hydrogel casting. *World J Urol* 2020;38:1631-41. DOI PubMed PMC
25. Golab A, Smektala T, Kaczmarek K, Stamirowski R, Hrab M, Słojewski M. Laparoscopic partial nephrectomy supported by training involving personalized silicone replica poured in three-dimensional printed casting mold. *J Laparoendosc Adv Surg Tech A* 2017;27:420-2. DOI
26. Maddox MM, Feibus A, Liu J, Wang J, Thomas R, Silberstein JL. 3D-printed soft-tissue physical models of renal malignancies for individualized surgical simulation: a feasibility study. *J Robot Surg* 2018;12:27-33. DOI
27. von Rundstedt FC, Scovell JM, Agrawal S, Zaneveld J, Link RE. Utility of patient-specific silicone renal models for planning and rehearsal of complex tumour resections prior to robot-assisted laparoscopic partial nephrectomy. *BJU Int* 2017;119:598-604. DOI PubMed PMC
28. Glybochko PV, Rapoport LM, Alyaev YG, et al. Multiple application of three-dimensional soft kidney models with localized kidney cancer: a pilot study. *Urologia* 2018;85:99-105. DOI
29. Hung AJ, Shah SH, Dalag L, Shin D, Gill IS. Development and validation of a novel robotic procedure specific simulation platform: partial nephrectomy. *J Urol* 2015;194:520-6. DOI PubMed
30. Makiyama K, Yamanaka H, Ueno D, et al. Validation of a patient-specific simulator for laparoscopic renal surgery. *Int J Urol* 2015;22:572-6. DOI PubMed
31. Centre for Evidence-Based Medicine. OCEBM levels of evidence. Available from: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebmllevels-of-evidence>. [Last accessed on 20 Nov 2023].
32. Yang B, Zeng Q, Yinghao S, et al. A novel training model for laparoscopic partial nephrectomy using porcine kidney. *J Endourol* 2009;23:2029-33. DOI
33. Ohtake S, Makiyama K, Yamashita D, Tatenuma T, Yamanaka H, Yao M. Validation of a kidney model made of N-composite gel as a training tool for laparoscopic partial nephrectomy. *Int J Urol* 2020;27:567-8. DOI PubMed
34. Gallagher AG, O'Sullivan GC. Fundamentals of surgical simulation. London: Springer; 2012. Available from: <https://link.springer.com/book/10.1007/978-0-85729-763-1>. [Last accessed on 20 Nov 2023].
35. Makiyama K, Tatenuma T, Ohtake S, Suzuki A, Muraoka K, Yao M. Clinical use of a patient-specific simulator for patients who were scheduled for robot-assisted laparoscopic partial nephrectomy. *Int J Urol* 2021;28:130-2. DOI PubMed
36. Kane MT. Validation. In: Brennan RL, editor. Educational measurement. 4th ed. Praeger; 2006. p. 17-64. Available from: <https://eric.ed.gov/?id=ED493398>. [Last accessed on 24 Nov 2023]
37. Salas E, Bowers CA, Rhodenizer L. It is not how much you have but how you use it: toward a rational use of simulation to support aviation training. *Int J Aviat Psychol* 1998;8:197-208. DOI PubMed
38. Satava RM. Virtual reality surgical simulator. The first steps. *Surg Endosc* 1993;7:203-5. DOI PubMed
39. Mazzone E, Puliatti S, Amato M, et al. A systematic review and meta-analysis on the impact of proficiency-based progression simulation training on performance outcomes. *Ann Surg* 2021;274:281-9. DOI
40. Maan ZN, Maan IN, Darzi AW, Aggarwal R. Systematic review of predictors of surgical performance. *Br J Surg* 2012;99:1610-21. DOI PubMed
41. Louangrath PI, Sutanapong C. Validity and reliability of survey scales. *Int J Res Methodol Soc Sci* 2018;4:99-114. DOI
42. Messick S. Validity. In: Linn RL, editor. Educational measurement. 3rd ed. American Council on Education and Macmillan; 1989. p. 13-104. Available from: <https://eric.ed.gov/?id=ED372105>. [Last accessed on 24 Nov 2023]
43. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull* 1955;52:281-302. DOI PubMed