BJS Open, 2021, zraa066
DOI: 10.1093/bjsopen/zraa066
Systematic Review

Virtual reality simulation in robot-assisted surgery: meta-analysis of skill transfer and predictability of skill

M. W. Schmidt[1], K. F. Köppinger[1], C. Fan[1], K.-F. Kowalewski[1,2], L. P. Schmidt[1], J. Vey[3], T. Proctor[3], P. Probst[1], V. V. Bintintan[4], B.-P. Müller-Stich[1] and F. Nickel[1,*]

[1]Department of General, Visceral, and Transplantation Surgery, University Hospital of Heidelberg, Heidelberg, Germany
[2]Department of Urology and Urological Surgery, University Medical Centre Mannheim, University of Heidelberg, Mannheim, Germany
[3]Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany
[4]Department of Surgery, First Surgical Clinic, University of Medicine and Pharmacy, Cluj Napoca, Romania

*Correspondence to: Department of General, Visceral, and Transplantation Surgery, University of Heidelberg, Im Neuenheimer Feld 420, 69120 Heidelberg, Germany (e-mail: felix.nickel@med.uni-heidelberg.de)

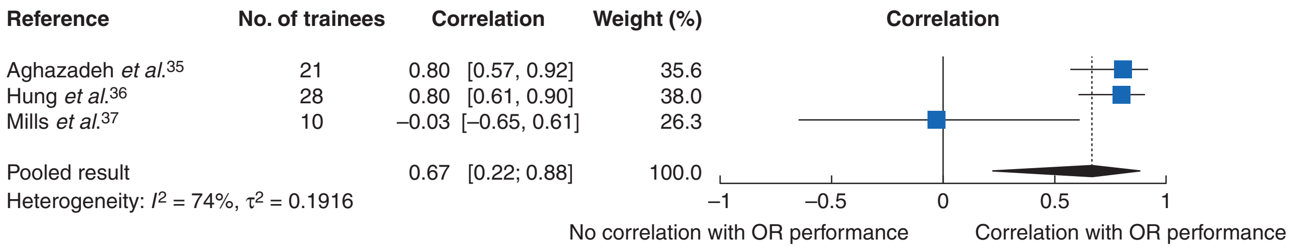
Abstract

Background: The value of virtual reality (VR) simulators for robot-assisted surgery (RAS) for skill assessment and training of surgeons has not been established. This systematic review and meta-analysis aimed to identify evidence on transferability of surgical skills acquired on robotic VR simulators to the operating room and the predictive value of robotic VR simulator performance for intraoperative performance.

Methods: MEDLINE, Cochrane Central Register of Controlled Trials, and Web of Science were searched systematically. Risk of bias was assessed using the Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale for Education. Correlation coefficients were chosen as effect measure and pooled using the inverse-variance weighting approach. A random-effects model was applied to estimate the summary effect.

Results: A total of 14 131 potential articles were identified; there were eight studies eligible for qualitative and three for quantitative analysis. Three of four studies demonstrated transfer of surgical skills from robotic VR simulators to the operating room measured by time and technical surgical performance. Two of three studies found significant positive correlations between robotic VR simulator performance and intraoperative technical surgical performance; quantitative analysis revealed a positive combined correlation ($r = 0.67$, 95 per cent c.i. 0.22 to 0.88).

Conclusion: Technical surgical skills acquired through robotic VR simulator training can be transferred to the operating room, and operating room performance seems to be predictable by robotic VR simulator performance. VR training can therefore be justified before operating on patients.



Graphical Abstract
This systematic review and meta-analysis presents current evidence on transferability of surgical skills acquired on robotic VR simulators to the real operating room, and on the predictability of intraoperative performance by robotic VR simulator performances. The limited data currently available support the use of robotic VR simulators for surgical skill acquisition and assessment.

Introduction

Robot-assisted surgery (RAS) is growing in popularity, with increasing numbers of procedures being undertaken in urology, gynaecology, and visceral surgery[1–3]. Although humans possess great flexibility and can adapt spontaneously to new situations in the operating room (OR), RAS brings the advantages of technology to improve precision and safety[4]. With three-dimensional vision, an ergonomic position at the console, tremor reduction, and no limitations on degrees of freedom of movement, RAS offers a multitude of benefits to the surgeon. Like any surgical modality, RAS requires appropriate and standardized training, which has yet to be achieved. Whether previous experience in open or laparoscopic surgery offers an advantage has not been determined[5–7]. Regardless of this, there are skills in RAS that need to be acquired by novice robotic surgeons, including

Received: September 29, 2020. **Accepted:** December 01, 2020

adjusting to lack of haptic feedback and fine motor handling. To make RAS training more accessible and shift training outside the OR, robotic virtual reality (VR) simulators have been developed. A wide range of courses offer VR simulator training, from basic to advanced skills, as well as operative procedures[8]. In laparoscopic surgery, VR simulators have been included in many training curricula[9–12] and have been validated widely[13,14]. There is growing evidence confirming skill transfer from laparoscopic VR simulators to the OR[15,16], although there is no compelling evidence for skill transfer from robotic VR simulators to the OR. Similarly, insufficient evidence exists regarding the predictability of OR surgical performance based on robotic VR simulator performance[17]. Evidence that skill transfer can be achieved could establish the role of robotic VR simulation within training curricula, and proving predictability of real-life surgical performance by robotic VR simulators would strengthen their role in the credentialing and selection of RAS surgeons. This systematic review aimed to present current evidence on skill transfer and prediction of skill between robotic VR simulation and real OR performance.

Methods

This systematic review was conducted in accordance with the PRISMA statement[18]. It was registered in the Prospective Register of Systematic Reviews (PROSPERO CRD42018111783).

Search strategy and information sources

MEDLINE (via PubMed), the Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science were searched[19], without restrictions on study design, language or publication date[19]. A librarian from Heidelberg University assisted in optimizing the search strategy based on the PICO criteria: P (patients/participants)—non-medical participants as novices, and medical students and doctors from operative specialties; I (intervention)—performance assessment and/or training on a robotic VR simulator plus OR assessment on human patients or live animal models; C (comparison)— no training, traditional RAS (da Vinci® Surgical System, Intuitive, Sunnyvale, California, USA)) training, comparison within intervention group plus OR assessment on human patients or live animal models; and O (outcome)—at least one measure of RAS operative performance/skill. All search strategies included terms related to RAS, robotic VR simulators, and skill assessment and transfer. Free-text words as well as index terms were used. An example of the search strategy in MEDLINE is provided in *Table S1*. The search was conducted on 5 June 2018. Studies included in the reference list of included articles or related systematic reviews and meta-analysis were screened for eligibility. Furthermore, studies that cited included articles or related systematic reviews or meta-analyses were identified using Google Scholar and screened for eligibility, even if published later than the abovementioned date. Grey literature was considered if enough data were provided and authors were contacted if necessary.

Eligibility criteria

Included studies were original articles that included transferability or predictability of surgical skill between robotic VR simulators and the OR (live animal model or human patients). The following studies were excluded: those involving surgical procedures other than thoracic, abdominal or pelvic surgery; those not providing a statistical assessment of skill transfer or predictability of skills; redundant patient populations; paediatric populations (aged less than 18 years); or failure to provide a full-text article. The original published protocol focused on assessing skill transfer to the OR, but during the screening process it became apparent that the terms skill transfer and predictability of robotic VR simulator to the OR were often mixed, misused or not specified. After reviewing the search strategies, which were broad enough to include all predictability studies, a decision was made to include the predictability of skills in this review. Previously screened abstracts were rescreened for the ability of surgical skills assessed on robotic VR simulators to predict skill in OR.

Outcomes

All types of RAS surgical skill assessments were included. Because of the heterogeneity in surgical skills assessment, all outcome parameters were categorized as time, technical surgical performance, operative outcome parameters, and patient-related outcome parameters (*Table 1*).

Study selection and data extraction

Title and abstract screening, as well as full-text screening and data extraction were performed by two authors independently. Disagreements were settled through discussion with a third author. Pretested standardized electronic spreadsheets were used for data extraction, and included an individual study identifier (author and year of publication), country, study population, study design, study process, result, key conclusions, and level of evidence according to the Oxford Centre for Evidence-Based Medicine[20].

Data synthesis and statistical analysis

Included studies were grouped by whether they assessed skill transfer from VR to the OR or the predictability of OR performance by robotic VR simulator performance (validity evidence). Owing to the broad inclusion criteria of this review, wide heterogeneity in the included studies was expected for study designs, participant types, and robotic VR simulators used. Studies assessing skill transfer were found to be too heterogeneous for a quantitative analysis, especially with regard to study design. The three studies that assessing the predictability of OR performance were found to be similar in design and were therefore included in a quantitative synthesis. As all three studies reported correlation coefficients, correlation was chosen as the effect measure for each study. As variance depends strongly on the correlation, the estimated effects were transformed to Fisher's z-scale before

Table 1 Classification of outcome parameters

	Definition	Example
Time	Time needed for procedure or task	Duration of operation
Technical surgical performance	Scores or parameters evaluating technical surgical performance, e.g., handling of instruments or efficiency	Objective Structured Assessment of Surgical Skills score, simulator metrics
Operative outcome parameters	Parameters assessing intraoperative outcome	Estimated blood loss, conversion rate
Patient-related outcome parameters	Postoperative patient-related outcomes	Length of stay, pain, complications

Table 2 Medical Education Research Study Quality Instrument, Newcastle–Ottawa Scale—Education scores, and risk of bias owing to funding for included studies

	MERSQI						NOS-E				Funding
	Study design	Sampling	Type of data	Validity	Data analysis	Outcome	Representativeness	Comparison group	Study retention	Blinding	Risk of bias
Skill transfer studies											
Culligan *et al.*[30]	■■□	■■□	■■■	■■■	■■■	■■□	□	■□□	■	■	Unclear
Gerull *et al.*[31]	■◧□	■■□	■■■	■■■	■■■	■■□	□	□□□	■	■	Low
Vargas *et al.*[32]	■■■	■■□	■■■	■■■	■■■	■◧□	■	■■■	■	■	Low
Wang *et al.*[33]	■■□	■■□	■■■	□□□	■■■	■■□	□	■■□	■	■	Low
Whitehurst *et al.*[34]	■■■	■■□	■■■	■■■	■■■	■◧□	□	■■■	■	■	Low
Predictability studies											
Aghazadeh *et al.*[35]	■□□	■■□	■■■	■■■	■■■	■■□					Low
Hung *et al.*[36]	■□□	■□□	■■■	■■■	■■■	■◧□					Low
Mills *et al.*[37]	■□□	■■□	■■■	■■■	■■■	■■□					Low

Each square represents a point that can be achieved in this category; more points equal better quality. ■, Full point given, ◧; half point given; □, no points given. MERSQI, Medical Education Research Study Quality Instrument; NOS-E, Newcastle–Ottawa Scale for Education.

analysis and its estimated variance was used for the synthesis. The summary effect and its confidence interval were converted back to correlations for presentation[21]. Owing to heterogeneity, a random-effects meta-analysis was used[22]. Inverse-variance weighting was used for combining the effect measures and the between-study variance τ^2 was estimated using the DerSimonian–Laird estimator[23]. The I^2 statistic was calculated to quantify statistical heterogeneity between the studies; 0–30 per cent represented no or only small, 30–60 per cent moderate, 60–90 per cent substantial, and 75–100 per cent considerable heterogeneity[24]. As only three studies were included in the meta-analysis, investigation of potential publication bias, sensitivity, and subgroup analysis were not addressed. The statistical analysis was performed using R version 3.6.3 with the meta package (R Project for Statistical Computing, Vienna, Austria)[25]. In accordance with Cohen[26], correlation coefficients equal to or greater than 0.1 were considered as small, those greater than 0.3 as medium, and those greater than 0.5 as large.

Risk-of-bias assessment

A modified Newcastle–Ottawa Scale for Education (NOS-E)[27,28] was used to assess the risk of bias of comparative studies (maximum 6 points). The Medical Education Research Study Quality Instrument (MERSQI) was used to assess methodological study quality[27,29], with a maximum of 18 points. Assessment categories for both tools are shown in *Table 2*. Assessment was undertaken by two authors independently and disagreements were settled in discussion with a third author. Judgements on MERSQI items were based on the definitions provided by Cook and Reed[27]. Points for validity were given if evidence of validity was cited. The choice of these two assessment tools was made after registering the protocol; because of the heterogeneity of study designs, the assessment tools chosen originally (Cochrane Collaboration tool for assessing risk of bias (RCT)[38] and Newcastle–Ottawa Scale[28]) were found not to be applicable to all studies. The NOS-E and MERSQI were chosen to complement each other as the NOS-E lacks items on objective assessment, validity evidence, data analysis, and level of outcomes, whereas the MERSQI lacks items on blinding and comparability. NOS-E was designed for comparative studies, and was not therefore used for non-comparative studies[27,29,39]. Funding of included studies is also included as a potential risk of bias, as this was not included in either score. All risk-of-bias assessment was done at a study level.

Results

A summary of the screening and selection process is shown in *Fig. 1*. Eight studies matched the inclusion criteria, five[30–34] assessing skill transfer from VR simulation to the OR (*Table 3*) and three[35–37] assessing the predictability of operative skills in the OR by robotic VR simulator performance (*Table 4*).

Evidence of skill transfer from robotic virtual reality simulators to the operating room

Included studies and study designs

A variety of designs were chosen for these five studies. Vargas and colleagues[32] and Wang *et al.*[33] undertook a RCT and NRCT. The intervention group was trained on a robotic VR simulator, whereas the control group received no further training, and both were compared on a procedure in the OR. Skill transfer was thought to have occurred if the intervention group outperformed the control group in the OR. A similar design was used by Whitehurst and co-workers[34], who also performed an RCT.

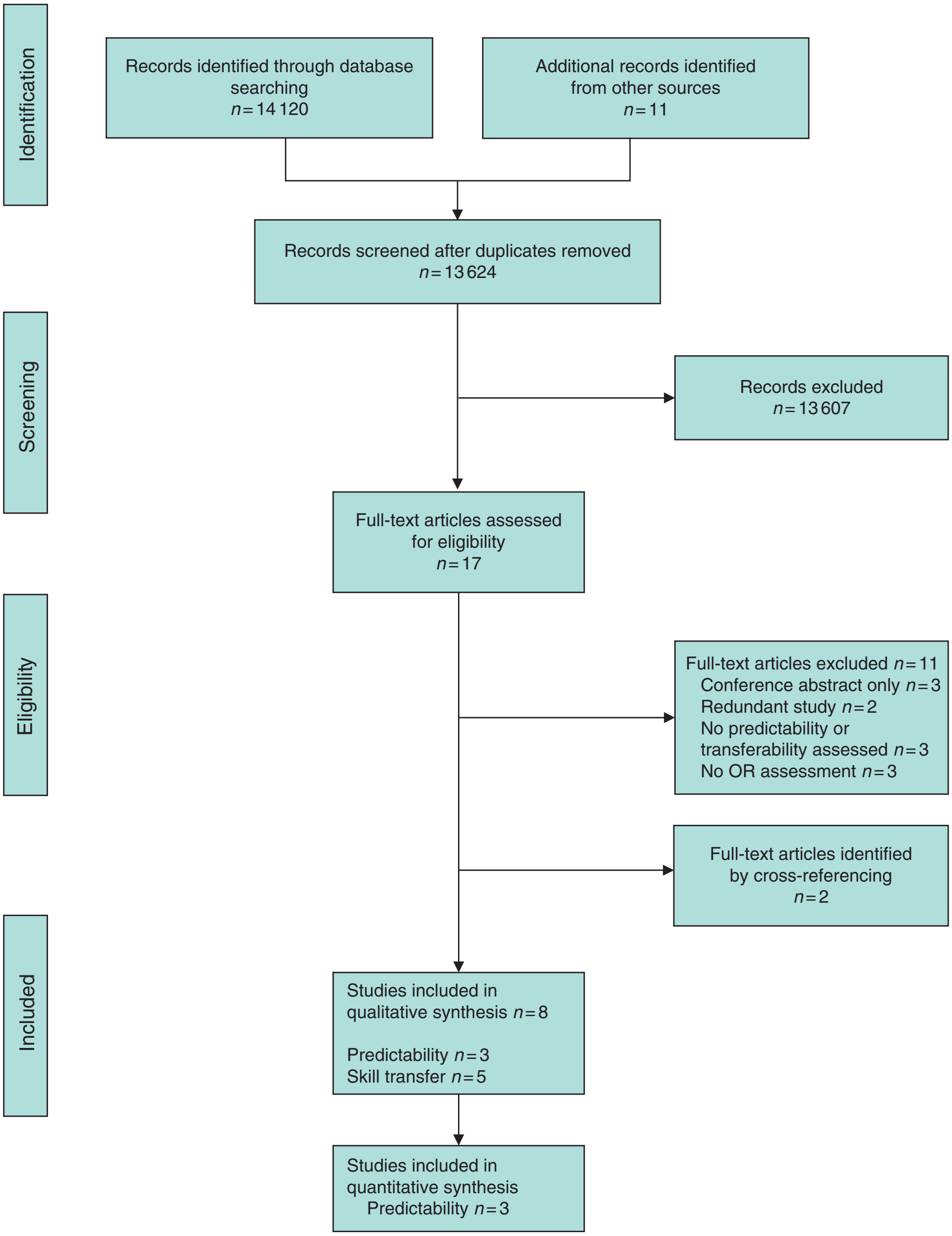


Fig. 1 PRISMA flow chart showing selection of articles for review

OR, operating room.

However, instead of the control group not receiving any training at all, their control group trained on the real robotic system to acquire RAS surgical skills. Therefore, skill transfer was thought to have occurred if the VR-trained group performed similarly to, or better than, the control group trained on the real robotic system in the live animal model post-test. This would indicate that the same or improved skills were acquired on the VR simulator as on the real robotic system, and that the skills acquired on the VR simulator could equally be transferred to the OR. Culligan and colleagues[30] recruited a RAS-naive intervention group and a RAS-credentialed control group for a non-randomized trial. The intervention group trained on the robotic VR simulator, whereas the control group received no further training. In this design, skill transfer was thought to have occurred if the intervention group performed better than, or equal to, the RAS-trained expert control group. Gerull *et al.*[31] opted not to use a control group. Instead, the intervention group performed a pretest and post-test on human patients, with robotic VR simulator training in between. Skill transfer was thought to have occurred if the post-test performance was significantly better than the pretest performance.

Participants

A total of 113 participants were assessed for skill transfer (*Table 3*). Surgical experience among participants varied widely. Most participants were RAS-naive and at varying stages of surgical training. Participants came from different specialties, including gynaecology (3 studies), urology (3 studies), and general surgery (1 study). One study[30] reported a predominantly female intervention group (11 of 14 participants) with a predominantly male control group

Table 3 Evidence of skill transfer from surgical skill acquired with robotic virtual reality simulators to the operating room

Reference Country	Design LOE	Simulator	No. of participants	Groups and experience	Intervention	Tasks		Outcomes	Results
						Simulator	OR		
Culligan *et al.*[30] USA	NRCT III	dVSS	18	IG: 14 credentialed gynaecological surgeons (naive to RAS) CG: 4 credentialed gynaecological surgeons (credentialed in RAS, but naive to dVSS simulator)	IG: online introduction, 10 tasks on dVSS until proficiency reached, standardized pig laboratory training, OR assessment CG: normal clinical activities, OR assessment	PB2, MB2, MB3, SS2, tubes, RW3, CT2, ED1, ED2, ES1	Robotic supracervical hysterectomy (human patients)	Operative time, EBL, GOALS	IG significantly outperformed CG in terms of operative time and EBL. No significant difference in mean GOALS scores
Gerull *et al.*[31] USA	NCT III	dVSS	31	Surgical residents naive to RAS (general surgery, urology, obstetrics and gynaecology)	Pretest/post-test test on live robotic procedure, in between completion of dVSS proficiency-based training curriculum	CT2, ED1, ES2, RR2, RW3, SS3, TR, tubes	Varying RAS procedures (human patients)	RO-SCORE, NTLX workload	Completion of dVSS curriculum associated with significant improvement across all domains of RO-SCORE and significant reduction of NTLX workload in all domains
Vargas *et al.*[32] USA	RCT II	dVSS	38	Medical students naive to RAS	IG: online introduction, baseline dVSS performance, 4 dVSS tasks to proficiency (maximum 10 ×), OR assessment CG: online introduction, baseline dVSS performance, no further training, OR assessment	CC1, SS1, SS2, tubes	Robotic cystostomy closure (live animal models)	GEARS, operating time	No significant differences between IG and CG
Wang *et al.*[33] China	NRCT III	dVSS	6	Certified robotic urologists, no robotic RARP experience	IG: baseline training on dVSS, 20 × tubes task on dVSS, OR assessment (9 patients per group) CG: no further training, OR assessment (9 patients per group)	Tubes	Robotic vesicourethral anastomosis (as part of RARP) (human patients)	Operating time (anastomosis and entire operation), EBL, creatinine in drainage, duration of catheter drainage, LOS	IG significantly faster than CG at creating anastomosis; no other differences between IG and CG
Whitehurst *et al.*[34] USA	RCT II	dV-Trainer	20	IG: 4 residents, 3 fellows, 3 attending surgeons (gynaecology and urology) naive to RAS CG: 2 residents, 6 fellows, 2 attending surgeons (gynaecology and urology) naive to RAS	IG: baseline cognitive skills and FLS test on dV, online didactic module, dV-Trainer tasks to proficiency, OR assessment CG: baseline cognitive skills and FLS test on dV; 3 FLS tasks on dV (PT, CC, ICSK) to proficiency, OR assessment	PP, RW1, PB1	Robotic cystostomy closure (live animal models)	GEARS, operating time, hand velocity	No significant differences between IG and CG on operative performance, which indicates skill transfer in this design

LOE, level of evidence according to the Oxford Centre for Evidence-Based Medicine; OR, operating room; NRCT, non-randomized controlled trial; dVSS, daVinci® Skills Simulator; IG, intervention group; CG, control group; RAS, robot-assisted surgery; PB, peg board; MB, match board; SS, suture sponge; RW, ring walk; CT, camera targeting; ED, energy dissection; ES, energy switcher; EBL, estimated blood loss; GOALS, Global Operative Ashtrtytty5r6sessment of Laparoscopic Skills; NCT, non-controlled trial; RR, ring and rail; TR, thread the rings; RO-SCORE, Robotic Ottawa Surgical Competency Operating Room Evaluation; NTLX, NASA Task Load Index; CC, camera clutching; GEARS, Global Evaluative Assessment of Robotic Skills; RARP, robot-assisted radical prostatectomy; LOS, length of stay; dV-Trainer, daVinci® Trainer; FLS, Fundamentals of Laparoscopic Surgery; dV, daVinci® Surgical System; PT, Peg Transfer; ICSK, Intracorporal suturing and knot tying.

Table 4 Evidence on the predictability of operative performance by robotic virtual reality simulator performances

Reference Country	Design LOE	Simulator	No. of participants	Groups and experience	Intervention	Tasks: Simulator	Tasks: OR	Outcomes	Results
Aghazadeh et al.[35] USA	Cross-sectional II	dVSS	21	17 urological trainees (residents; 0–55 RAS procedures performed) 4 urological RAS experts (fellow/attendings; 58–600 RAS procedures performed)	Instructional videos, completion of 8 dVSS tasks followed by OR assessment	PB1, PB2, RR2, RW3, MB3, SS3, tubes, ES	Robotic endopelvic fascia dissection (part of RARP procedure) (human patients)	dVSS: simulator score dV: GEARS	Good correlation between dVSS simulator scores and GEARS scores including subdomains, except for dVSS exercise ES1
Hung et al.[36] USA	Cross-sectional II	Modified dV-Trainer	28	Expert surgeons (105–3000 RAS procedures performed) Intermediate surgeons (0–75 RAS procedures performed)	Assessment on dV-Trainer followed by OR assessment	Renorrhaphy	Robotic RPN (live porcine model)	dVSS: GEARS dV: GEARS	High correlation between GEARS scores for VR renorrhaphy and GEARS scores for RPN on live animal model for total GEARS scores and each subdomain
Mills et al.[37] USA	Cross-sectional II	dVSS	10	Attending robotic surgeons from gynecology (4), urology (4), thoracic surgery (1), and general surgery (1). 20–346 RAS operations in past 4 years.	Completion of 4 dVSS tasks followed by 2 OR assessments	CT1, RW3, SS3, ED3	Next scheduled RAS operation of each surgeon (human patients)	dVSS: simulator score dV: GEARS	No correlation between dVSS simulator scores and intraoperative GEARS scores

LOE, level of evidence according to the Oxford Centre for Evidence-Based Medicine; OR, operating room; dVSS, daVinci® Skills Simulator; RAS, robot-assisted surgery; PB, peg board; RR, ring and rail; RW, ring walk; MB, match board; SS, suture sponge; ES, energy switcher; RARP, robot-assisted radical prostatectomy; dV, daVinci® Surgical System; GEARS, Global Evaluative Assessment of Robotic Skills; dV-Trainer, daVinci® Trainer; RPN, robotic partial nephrectomy; VR, virtual reality; CT, camera targeting; ED, energy dissection.

(1 of 4 participants). Vargas and colleagues[32] assessed a sex-balanced group with a total of 19 women and 19 men, evenly distributed between the intervention and control groups. The remaining three studies did not provide data on the sex of the participants.

Tasks and operative procedures

A total of 19 different tasks were used during robotic VR simulator training, most commonly the tubes task (4 studies), a suturing and knot tying task (*Table 3*). Wang and co-workers[33] only assessed the tubes task, whereas other studies chose a variety of tasks (ranging from 3 to 10 per study) for a broader skill spectrum assessment, including camera tasks, object transfer tasks, and needle manipulation tasks with varying levels of difficulty.

Clinical operative skills were assessed on human patients in three studies[30,31,33] and on live animal models in two[32,34]. Only two studies evaluated whole RAS procedures. Procedures in the study of Gerull *et al.*[31] varied from participant to participant, as well as from pretest to post-test of each participant. Culligan and co-workers[30] assessed all participants on the same procedure, a robotic-assisted supracervical hysterectomy. Wang and colleagues[33] assessed participants on the creation of a vesicourethral anastomosis as part of a robotic-assisted radical prostatectomy, similar to the VR tubes task that participants in their intervention group practised on. Whitehurst *et al.*[34] and Vargas and co-workers[32] chose a more simplified procedure (cystostomy closure) for assessment.

Outcomes

Outcome measures assessing surgical skill focused on four different aspects: technical surgical performance, time, operative outcomes, and patient-related outcomes. Four studies[30–32,34] evaluated technical surgical performance using the highly validated scoring systems Global Evaluative Assessment of Robotic Skills (GEARS)[40], Global Operative Assessment of Laparoscopic Skills (GOALS)[41], and the Robotic Ottawa Surgical Competency Operating Room Evaluation (RO-SCORE), which was adapted by the authors from the validated O-SCORE[42], to better match RAS procedures. GEARS and the RO-SCORE are specific for RAS procedures, whereas GOALS was originally developed for laparoscopic procedures. Time was assessed in four studies[30,32–34], and operative outcome measures (estimated blood loss) in two[30,33]. Only one study[33] examined patient-related outcomes, including duration of catheter drainage and length of stay. Although not technically assessing surgical skill, Gerull and co-workers[31] also looked at changes in self-rated workload using the NASA Task Load Index after training on the robotic VR simulator.

Findings

Table 5 shows a summary of outcomes for which skill transfer was demonstrated. Three of four studies showed skill transfer with regard to surgical technical performance and time. One of two studies indicated skill transfer for operative outcome parameters, and the only study assessing patient-related outcomes did not show skill transfer. For the only non-objective outcome parameter, Gerull and colleagues[31] reported significantly reduced workload in the intervention group after training with the robotic VR simulator in all domains (mental demand, physical demand, temporal demand, performance, effort, and frustration).

Table 5 Summary of results of surgical skill transfer assessment by outcome

Reference	Surgical technical performance		Time	Operative outcome parameters	Patient-related outcome parameters
	Measure	Score	Operating time (min)	Blood loss (ml)	Length of stay (days)
Culligan et al.[30]	GOALS	✓	✓	✓	
Intervention group		34.7	21.7(3.3)	25.4	
Control group		31.1 +	30.9 (0.6)*	31.3*	
Gerull et al.[31]	RO-SCORE	✓			
Pretest		2.06(0.85)			
Post-test		4.35(0.69)†			
Vargas et al.[32]	GEARS	✗	✗		
Intervention group		15.4(2.5)	9.2(2.7)		
Control		15.3(3.4)	9.9(2.1)		
Wang et al.[33]			✓	✗	✗
Intervention group			25.1 (7.1)	130.0(55.2)	3.6(1.1)
Control group			40.0(12.4)*	121.1(40.1)	4.2(1.0)
Whitehurst et al.[34]	GEARS	✓	✓		
Simulator		2.83(0.66)	NS†		
Real robot		2.96(0.77)†			

Values are mean(s.d.). GOALS, Global Operative Assessment of Laparoscopic Skills; RO-SCORE, Robotic Ottawa Surgical Competency Operating Room Evaluation; GEARS, Global Evaluative Assessment of Robotic Skills. ✓, Evidence of skill transfer; ✗, no evidence of skill transfer. *$P < 0.050$. †No significant difference (NS); indicates skill transfer in this study design). + No significant difference; skill transfer is indicated by equal or better performance in this study design.

Prediction of operative performance

Included studies and study designs

Three studies[35–37] were cross-sectional evaluations to determine the predictability of operative performance by robotic VR simulator performances. Participants with varying degrees of RAS experience undertook a number of tasks on the robotic VR simulator before intraoperative assessment of their surgical performance using the real robotic system (2 studies on human patients, 1 live porcine model). Outcome parameters were then correlated to assess whether VR performances could predict operative performances. All studies assessed OR performance at the same time as VR performance (concurrent validity evidence) (*Table 4*).

Participants

A total of 59 participants were assessed. One study[35] included participants from a single specialty (urology), one[37] included attending RAS surgeons from four different specialties, and the third did not state the surgical specialty of participants[36]. The previous RAS experience of participants varied between studies, and only one[35] reported the sex of participants.

Tasks and operative procedures

Ten different tasks were assessed on the robotic VR simulator by Aghazadeh and colleagues[35] and Mills *et al.*[37] (*Table 4*). Although most exercises were designed to train a specific basic skill such as bimanual dexterity or use of electrocautery, the tasks suture sponge, tubes, and renorrhaphy focused on suturing and knot-tying.

One study[35] assessed all participants on endopelvic fascia dissection with the real robotic system as part of a robotic-assisted radical prostatectomy procedure. Only one patient was taken into account for analysis. Mills and co-workers[37] evaluated two consecutive surgical procedures. However, procedures varied as the next two scheduled operations for each surgeon were assessed (no further details on type of procedures were available). Only Hung and colleagues[36] used a live porcine model instead of human patients to assess intraoperative performance on a robotic-assisted partial nephrectomy.

Outcomes

Two studies[35,37] used the simulators' built-in composite score (simulator score) based on a number of single metrics, such as time, path length, and economy of motion. Hung and colleagues[36] used GEARS to assess robotic VR renorrhaphy rather than simulator metrics. Clinical technical surgical performance was assessed with the help of GEARS in all three studies as it considers depth perception, bimanual dexterity, efficiency, force sensitivity, autonomy, and robotic control[35]. Mills and co-workers[37] focused on the total GEARS score only (excluding autonomy), whereas Aghazadeh *et al.*[35] and Hung and colleagues[36] assessed each domain of GEARS separately as well as the total GEARS score.

Findings

Random-effects meta-analysis based on Fisher's z-transformation of correlation coefficients revealed a positive pooled correlation ($r = 0.67$, 95 per cent c.i. 0.22 to 0.88) between robotic VR simulator performance (combined over all tasks assessed) and intraoperative performance (*Fig. 2*). There was substantial heterogeneity between the studies. Aghazadeh and colleagues[35] showed a good correlation between robotic simulator scores and total GEARS scores for all tasks ($r = 0.582$–0.784, $P < 0.050$) except energy switcher, for which the correlation was not statistically significant ($r = 0.412$, $P = 0.063$). Further analysis revealed that different tasks correlated better with certain GEARS domains. For example, suturing exercises (suture sponge 3 and tubes) as well as exercises moving objects in space (ring and rail 2) correlated best with bimanual dexterity ($r = 0.716$–0.763; $P \leq 0.001$), whereas pick and place tasks (peg board 1 and 2, match board 3) and ring walk 3 correlated best with depth perception ($r = 0.675$–0.810, $P \leq 0.003$). Overall, autonomy correlated best with most exercises. Similarly, Hung and co-workers[36] reported a high correlation between the total GEARS scores for the robotic VR simulated task and the total GEARS score of a robotic-assisted partial nephrectomy in a live porcine model ($r = 0.8$, $p < 0.0001$), as well as between all GEARS subdomains ($r = 0.7$–0.9, $P < 0.001$).

Risk of bias and study quality

Studies assessing the predictability of operative skill by robotic VR simulator performances had a mean±SD MERSQI score of 13.5±0.7 (median 14, range 12.5–14). Skill transfer studies achieved a mean score of 14.5±1.3 (median 15, range 12–15.5). The skill transfer studies received a mean NOS-E score of 4±1.4

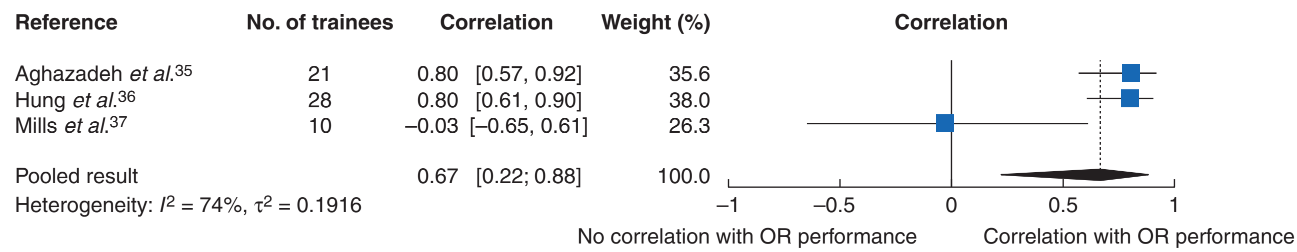


Fig. 2 Forest plot showing correlation between robotic virtual reality simulator performance and operating room performance

A random-effects model was used for meta-analysis. Correlations are shown with 95 per cent confidence intervals. OR, operating room.

(median 4, range 2–6). Only one RCT[32] achieved a score of 6. In general, points were often not given owing to incomplete reporting, especially information regarding the representativeness of the intervention group (*Table 1*).

Discussion

This review has shown that certain surgical skills acquired on robotic VR simulators can be transferred to the OR and that technical performance in the OR seems to be predictable by robotic VR simulator performance. The extent to which this can be applied, however, remains unclear.

Determining whether surgical skills acquired on a robotic VR simulator can be transferred to an actual operation is crucial to strengthen the role of VR robotic simulators in RAS skill acquisition. Although skill transfer studies need to be distinguished from studies assessing the predictability of intraoperative performance (concurrent or predictive validity), the terminology associated with validity is often misunderstood and misused. In this systematic review, five studies were identified that assessed the transferability of surgical skills. They proved to be heterogeneous with regards to participants, study design, tasks, and outcome. Two main issues with the study design became evident that might have influenced the assessment of surgical skill transfer: the training periods on the robotic VR simulators were insufficient to acquire significant technical skill; and operative performance assessment did not adequately measure the skill set acquired by training with the robotic VR simulator. Well planned, high-quality studies are thus necessary to minimize the risk of bias in skill transfer studies.

Shortened operating times as demonstrated in this review could justify the role of robotic VR simulators in RAS training, as time in the OR is expensive. A current analysis reported a cost of US $37 (€31; exchange rate 21 February 2021) per minute of OR time[43], although a recent international Delphi survey[44] reported that laparoscopic surgery experts considered time to be the least important indicator of good surgical performance on VR simulators (compared with safety, dexterity, and efficiency) . Exactly how best to implement robotic VR simulators in training curricula (such as duration of training, choice of tasks) requires further investigation[45]. In addition, to fully understand the benefits and limitations of robotic VR simulator training, non-technical aspects of skill should be assessed for skill transfer, such as cognitive training[46] and clinical decision-making.

Different concepts of validity are still frequently used to assess new scoring systems or training modalities, and variation in terminology and practices is common. A currently accepted and widespread concept of validity is based on Messick's framework, which conceptualizes all validity under one overarching framework of construct validity[47–49]. One aspect of validity evidence within this framework includes answering the following questions: how accurately do robotic VR simulator performances predict current intraoperative performance (concurrent validity), and how accurately do they predict future intraoperative performance (predictive validity)? Answering the first question can help define the role of robotic VR simulators as an assessment method for credentialing, whereas the second question might help with the selection of future robotic surgeons. In the present review, evidence to answer these questions was rare; not a single study comprehensively evaluated the predictability of future operative performance. The meta-analysis of studies of predictability revealed a positive pooled correlation between robotic VR simulator performance (combined over all tasks assessed) and intraoperative performance, but had a broad confidence interval indicating high degree of uncertainty with regard to the point estimate of the summary effect. Because of the limited number of studies and the known challenges in estimating between-study heterogeneity in such settings, the width of the confidence interval may have been underestimated.

Assessing skill transfer in relation to clinical operative performance is complex. Operative procedures cannot be standardized like laboratory tasks. Procedural difficulty varies with patient-specific characteristics, such as anatomy, general condition, previous surgical and medical history, and co-morbidities. The extent of pretraining, the types of task in relation to the operative performance assessed, and the amount of training and assessment tools used to evaluate surgical skill, are all important variables. As summarized in *Tables 3 and 4*, study designs, tasks, procedures, and participants of studies included in the present review were heterogeneous. Results cannot simply be attributed to robotic VR simulators alone but must be considered in context of the study design. Systematic reviews evaluating skill transfer from laparoscopic VR trainers have identified similar issues with variability between studies[16]. Although all studies included in this review focused on technical aspects of surgical skill, only one[31] additionally showed a reduction in mental workload as a result of training on VR simulators, and an improvement in communication/use of staff, as a subdomain of the RO-SCORE. No study specifically assessed clinical decision-making or communication with validated tools; these are crucial skills for expert surgeons. Most studies were limited by their small sample size. Only two studies[31,32] assessed more than 30 participants. Risk-of-bias assessments revealed that the studies demonstrated good methodological quality. NOS-E and MERSQI scores were quite homogeneous and generally high, so it seems unlikely that differences in study quality influenced the results greatly.

A systematic review in 2016 reported evidence of validity and skill transfer of robotic VR simulators, based on the historical terminology of validity, although most of the evidence did not come from assessment of live animal models or human patients[50]. It was concluded that there was no evidence of skill transfer from simulation to clinical surgery on patients. A further review[51] in 2016 included only one study on human patients, claiming to assess predictive validity of robotic VR simulators. With a total of eight live animal and human patient studies, the present systematic review and meta-analysis provides the collective evidence that surgical skill acquired on robotic VR simulators is transferable to the OR with regard to time and technical surgical performance. Performance on robotic VR simulators seems to predict current technical RAS performance in the OR. These data suggest that there are potential benefits that could justify the use of robotic VR simulators in dedicated training curricula, and emphasize the values imulation training before performance in the real OR to ensure patient safety.

Funding

No funding was received for this study.

Acknowledgements

The authors thank D. Fleischer from Heidelberg University Library for support in conducting the literature search.

Disclosure. F.N. reports receiving travel support for conference participation as well as equipment provided for laparoscopic surgery courses by Karl Storz, Johnson & Johnson, Intuitive Surgical, Cambridge Medical Robotics, and Medtronic. K.-F.K. received travel and non-financial support from Intuitive Surgical and CMR Surgical. V.V.B. reports tutoring surgical training programmes sponsored by Medtronic, being an invited speaker for Covidien, Bayer, Abbvie, and Karl Storz, and receiving support for congress participation from Johnson & Johnson. The authors declare no other conflict of interest.

Supplementary material

Supplementary material is available at *BJS Open* online.

References

1. Sood A, Jeong W, Peabody JO, Hemal AK, Menon M. Robot-assisted radical prostatectomy: inching toward gold standard. *Urol Clin North Am* 2014;**41**:473–484
2. Diana M, Marescaux J. Robotic surgery. *Br J Surg* 2015;**102**: e15–e28
3. Intuitive Surgical. Annual Report 2018. http://www.annualreports.com/Company/intuitive-surgical-inc (accessed 20 February 2020)
4. Zhou NX, Chen JZ, Liu Q, Zhang X, Wang Z, Ren S *et al.* Outcomes of pancreatoduodenectomy with robotic surgery *versus* open surgery. *Int J Med Robotics Comput Assist Surg* 2011;**7**: 131–137
5. Kowalewski KF, Schmidt MW, Proctor T, Pohl M, Wennberg E, Karadza E *et al.* Skills in minimally invasive and open surgery show limited transferability to robotic surgery: results from a prospective study. *Surg Endosc* 2018;**32**:1656–1667
6. Cumpanas AA, Bardan R, Ferician OC, Latcu SC, Duta C, Lazar FO. Does previous open surgical experience have any influence on robotic surgery simulation exercises? *Wideochir Inne Tech Maloinwazyjne* 2017;**4**:366–371
7. Angell J, Gomez MS, Baig MM, Abaza R. Contribution of laparoscopic training to robotic proficiency. *J Endourol* 2013;**27**: 1027–1031
8. Julian D, Tanaka A, Mattingly P, Truong M, Perez M, Smith R. A comparative analysis and guide to virtual reality robotic surgical simulators. *Int J Med Robot* 2017;**14**:e1874
9. Nickel F, Brzoska JA, Gondan M, Rangnick HM, Chu J, Kenngott HG *et al.* Virtual reality training *versus* blended learning of laparoscopic cholecystectomy: a randomized controlled trial with laparoscopic novices. *Medicine (Baltimore)* 2015;**94**:e764
10. Kowalewski KF, Minassian A, Hendrie JD, Benner L, Preukschas AA, Kenngott HG *et al.* One or two trainees per workplace for laparoscopic surgery training courses: results from a randomized controlled trial. *Surg Endosc* 2018;**33**:1523–1531
11. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G *et al.* Virtual reality simulation for the operating room. *Ann Surg* 2005;**241**:364–372
12. Nickel F, Jede F, Minassian A, Gondan M, Hendrie JD, Gehrig T *et al.* One or two trainees per workplace in a structured multimodality training curriculum for laparoscopic surgery? Study protocol for a randomized controlled trial—DRKS00004675. *Trials* 2014;**15**:137
13. Thijssen AS, Schijven MP. Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? *Am J Surg* 2010;**199**:529–541
14. Yiannakopoulou E, Nikiteas N, Perrea D, Tsigris C. Virtual reality simulators and training in laparoscopic surgery. *Int J Surg* 2015; **13**:60–64
15. Seymour NE. VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. *World J Surg* 2008;**32**:182–188
16. Dawe SR, Pena GN, Windsor JA, Broeders JA, Cregan PC, Hewett PJ *et al.* Systematic review of skills transfer after surgical simulation-based training. *Br J Surg* 2014;**101**:1063–1076
17. American Educational Research Association, American Pschyological Association, National Council on Measurement in Education (eds). *Standards for Educational and Psychological Testing*. American Educational Research Association. Washington. 2014
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;**151**:264–269
19. Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Buchler MW *et al.* Optimal literature search for systematic reviews in surgery. *Langenbecks Arch Surg* 2018;**403**:119–129
20. Oxford Levels of Evidence Working Group. *The Oxford 2011 Levels of Evidence*. Oxford: Oxford Centre for Evidence-Based Medicine, 2011
21. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Effect sizes based on means. *Introduction to Meta-Analysis* 2009;21–32.
22. Borenstein M, Hedges L, Rothstein H, Meta-analysis: fixed effect vs. random effects. Meta-analysis com. 2007.
23. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–188
24. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M *et al.* *Cochrane Handbook for Systematic Reviews of Interventions version 6.0 [Updated* July 2019]. www.training.cochrane.org/handbook (accessed 29 January 2020)
25. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Mental Health* 2019;**22**:153–160
26. Cohen H. *Statistical Power Analysis for Behavioral Sciences*. Hillsdale: Lawrence Erlbaum Associates, 1988
27. Cook DA, Reed DA. Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale—Education. *Acad Med* 2015;**90**:1067–1076

28. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M *et al.* *The Newcastle–Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses*. Ottawa: Ottawa Hospital Research Institute, 2014
29. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA* 2007;**298**:1002–1009
30. Culligan P, Gurshumov E, Lewis C, Priestley J, Komar J, Salamon C. Predictive validity of a training protocol using a robotic surgery simulator. *Female Pelvic Med Reconstr Surg* 2014;**20**:48–51
31. Gerull W, Zihni A, Awad M. Operative performance outcomes of a simulator-based robotic surgical skills curriculum. *Surg Endosc* 2020;**34**:4543–4548
32. Vargas MV, Moawad G, Denny K, Happ L, Misa NY, Margulies S *et al.* Transferability of virtual reality, simulation-based, robotic suturing skills to a live porcine model in novice surgeons: a single-blind randomized controlled trial. *J Minim Invasive Gynecol* 2017;**24**:420–425
33. Wang F, Zhang C, Guo F, Sheng X, Ji J, Xu Y *et al.* The application of virtual reality training for anastomosis during robot-assisted radical prostatectomy. *Asian J Urol* 2019, doi: 10.1016/j.ajur.2019.11.00
34. 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). *J Minim Invasive Gynecol* 2015;**22**:483–488
35. Aghazadeh MA, Mercado MA, Pan MM, Miles BJ, Goh AC. Performance of robotic simulated skills tasks is positively associated with clinical robotic surgical performance. *BJU Int* 2016; **118**:475–481
36. 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**:520–526
37. Mills JT, Hougen HY, Bitner D, Krupski TL, Schenkman NS. Does robotic surgical simulator performance correlate with surgical skill? *J Surg Educ* 2017;**74**:1052–1056
38. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;**366**:I4898
39. Reed DA, Beckman TJ, Wright SM, Levine RB, Kern DE, Cook DA. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. *J Gen Intern Med* 2008;**23**:903–907
40. 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**:247–252
41. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondre K, Stanbridge D *et al.* A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 2005;**190**:107–113
42. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med* 2012; **87**:1401–1407
43. Childers CP, Maggard-Gibbons M. Understanding costs of care in the operating room. *JAMA Surg* 2018;**153**:e176233
44. Schmidt MW, Kowalewski KF, Schmidt ML, Wennberg E, Garrow CR, Paik S *et al.* The Heidelberg VR Score: development and validation of a composite score for laparoscopic virtual reality training. *Surg Endosc* 2019;**33**:2093–2103
45. Kowalewski KF, Garrow CR, Proctor T, Preukschas AA, Friedrich M, Muller PC *et al.* LapTrain: multi-modality training curriculum for laparoscopic cholecystectomy—results of a randomized controlled trial. *Surg Endosc* 2018;**32**:3830–3838
46. Kowalewski KF, Hendrie JD, Schmidt MW, Proctor T, Paul S, Garrow CR *et al.* Validation of the mobile serious game application Touch Surgery^{TM} for cognitive training and assessment of laparoscopic cholecystectomy. *Surg Endosc* 2017;**31**:4058– 4059
47. Van Nortwick SS, Lendvay TS, Jensen AR, Wright AS, Horvath KD, Kim S. Methodologies for establishing validity in surgical simulation studies. *Surgery* 2010;**147**:622–630
48. Messick S. Educational measurement. In: Linn R (ed.), *Validity*. (3rd edn). New York, NY: American Council on education and Macmillan, **1989**, 13–103
49. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol* 1995;**50**:741–749
50. Moglia A, Ferrari V, Morelli L, Ferrari M, Mosca F, Cuschieri A. A systematic review of virtual reality simulators for Robot-assisted Surgery. *European Urology* 2016;**69**:1065–1080. doi: 10.1016/j.eururo.2015.09.021.
51. Bric JD, Lumbard DC, Frelich MJ, Gould JC. Current state of virtual reality simulation in robotic surgery training: a review. *Surg Endosc* 2016;**30**:2169–2178