Surgical Endoscopy
https://doi.org/10.1007/s00464-022-09593-0

REVIEW ARTICLE

Training and assessment using the LapSim laparoscopic simulator: a scoping review of validity evidence

Conor Toale[1] · Marie Morris[1] · Dara O. Kavanagh[1]

Received: 27 April 2022 / Accepted: 25 August 2022

Abstract

Introduction The LapSim (Surgical Science, Sweden) laparoscopic simulator is a high-fidelity virtual reality simulator for use in endoscopic surgical training. This review critiques the current validity evidence for the LapSim laparoscopic simulator, specifically with respect to its potential use as a tool and method of training and assessment in surgery.

Methods A scoping review of the MEDLINE (PubMed), EMBASE, Cochrane and Web of Science databases was conducted in accordance with PRISMA guidelines (2020)—scoping review extension. Articles were included if they presented validity evidence for the use of the LapSim in operative skill training or assessment, in accordance with Messick's validity framework. European Association of Endoscopic Surgeons (EAES) guidelines (2005) were used to provide recommendations for the use of the LapSim in operative performance training and assessments.

Results Forty-nine articles were included. An EAES level 2 recommendation was provided with regard to the internal consistency reliability of automated performance metrics in assessing performance. An EAES recommendation of 2 was awarded with respect to the ability of the LapSim to discriminate based on case volume and overall laparoscopic experience (relationships with other variables). Performance assessment metrics on the LapSim correlate with improved performance in the operating room (EAES level of recommendation 1—consequential validity).

Conclusion The LapSim has accumulated substantial evidence supporting the validity of its use in surgical training and assessment. Future studies should explore the relationship between the achievement of performance benchmarks on the LapSim and subsequent patient outcomes, and interrogate the benefits of implementing virtual reality simulation training and assessment curricula in post-graduate surgical training programmes.

Keywords Surgery · Simulation · Assessment · Technical skill · Competency-based education · Operative performance · Virtual reality

Simulation is becoming an established tool in surgical education and training [1]. Its role in assessment, either for surgeon selection, credentialing, certification or re-certification, is less well explored [2]. The validity of any simulation-based assessment framework relates not only to the tools used, but also the simulation models, be they 'low-fidelity' bench models or 'high-fidelity' virtual reality platforms [3]. In 2005, the Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes, as a subgroup of the European Association of Endoscopic Surgeons (EAES), produced consensus guidelines for the validation of virtual reality simulators [4]. Using a classical validity framework, the evidence supporting the validity of available simulators was graded. The LapSim (Surgical Science, Sweden) system achieved the lowest level of recommendation (level 4) for face, content, construct and concurrent validity for all of its training or assessment modules [4].

Fairhurst et al. in 2010 conducted a further review of published validity evidence for the LapSim system [5]. Since 2010, a PubMed search using the term 'LapSim' retrieves 51 published studies in relation to this simulator. The aim of this review is therefore to collate and critique updated validity evidence for the LapSim laparoscopic simulator, specifically with respect to its potential use in the training and assessment of surgical skill. We sought to present this evidence in accordance with a contemporary validity

✉ Conor Toale
conortoale@rcsi.com

1 Department of Surgical Affairs, Royal College of Surgeons in Ireland, 121 St Stephen's Green, Dublin 2 D02 YN77, Ireland

Published online: 19 September 2022

framework (Messick's), and provide recommendation consistent with levels of evidence and recommendations outlined by the 2005 EAES guidelines.

Methods

This scoping review is reported in accordance with the Preferred Items for the Reporting of Systematic Reviews and Meta-Analyses (PRISMA-ScR) guidelines 2018 (Scoping Review extension) [6]. A scoping review was deemed to be the most appropriate search strategy given the aim of this review; namely to map and critique the available validity evidence for the LapSim. The number of anticipated outcome measures across validity domains, along with the expected heterogeneity of eligible studies across populations and specialties, favoured a scoping rather than a systematic review. Eligibility criteria are outlined in Table 1.

The MEDLINE (PubMed), EMBASE, Cochrane and Web of Science databases were queried from database inception to present (last search March 3rd 2022) using the search term 'LapSim'. No limits or filters were applied. The reference lists of included full texts were also searched for additional articles, screened according to the same inclusion and exclusion criteria outlined in Table 1. Duplicate results were removed. Titles, abstracts and key words were screened for relevance. Full text articles were screened by a single reviewer (CT). Where uncertainty regarding the inclusion of an article existed, a consensus decision was reached by the research team (CT, MM, DK). Data from each included article was extracted using pre-defined pro-forma (Appendix 1, supplemental material), in accordance with Messick's validity evidence framework (Table 2). The quality and level of recommendation for relevant data was assessed according to the EAES guidelines for the evaluation of validity evidence related to virtual reality simulators (Table 3). The quality of included studies was assessed using the Medical Education Research Study Quality Instrument (MERSQI) [7]. Evidence was tabulated and further described by narrative review, according to the five sources of validity evidence outlined in Table 2.

Results

From an initial yield of 412 results, 42 eligible studies were identified (Fig. 1). A further seven eligible articles were identified from a manual search of reference lists. Therefore, 49 articles in total were included in a narrative synthesis (Fig. 1). Study characteristics are outlined in Table 4. The mean Medical Education Research Study Quality Instrument (MERSQI) score of included studies was 13.5, range 11 – 15.5 (supplemental data).

Content validity

Eleven included studies report relevant content validity data for the LapSim across simulated tasks in gynaecological, cardiothoracic and general surgery, as well as basic endoscopic skill assessment [11–13, 15, 17, 18, 27, 29, 33, 37, 44]. A summary of published content validity evidence is outlined in Table 5.

Only two included studies defined appropriate procedural tasks for training or assessment using consensus methodology. Schreuder et al. [37] and van Dongen et al. defined tasks appropriate for assessment of colorectal skills and basic surgical skills respectively [33]. Data pertaining to the development and selection of relevant automated metrics was lacking across included studies. Given the inconsistent findings across studies of disparate methodologies, and a

Table 1 Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Randomised controlled trials, cohort studies, case–control studies, other quasi-experimental studies	Case reports, technique articles, editorials, letters to the editor
Data pertaining to validity as categorised by Messick's validity framework, namely content validity, internal structure validity, response process validity, relationships with other variables and consequences	Data not relating to validity as outlined by Messick's framework
Performance data: performance as assessed using the LapSim system, either according to automated metrics or objective assessment by a trained assessor observing performance on the LapSim	Performance data: data generated using any other laparoscopic simulator, bench-top models, or other simulation models
Survey or qualitative data pertaining to the LapSim system	Survey or qualitative data relating to other laparoscopic simulators, bench top models, or other simulation models
Data quantifying the relationship between performance as measured on or using the LapSim and other variables of interest relevant to operative performance or skill	Data pertaining to the relationship between LapSim measurement data and other variables not directly related to operative performance, such as hand-eye co-ordination, baseline psychometric measurements or non-surgical technical ability such as musical instrument playing, video gaming ability or sporting ability

Table 2 The five sources of evidence according to Messick's validity framework. Adapted from [10]

Source of evidence	Definition	Examples of evidence
Content	"The relationship between the content of a test and the construct it is intended to measure" [8]	Procedures for item sampling and scoring (e.g. expert panel, previously described instrument, test blueprint and pilot testing and revision)
Internal Structure	Relationship amongst data items within the assessment and how these relate to the overarching construct	Internal consistency reliability, interrater reliability, factor analysis, test item statistics
Relationships with other variables	"Degree to which these relationships are consistent with the construct underlying the proposed test score interpretations" [8]	Correlation with tests measuring similar constructs. Correlation (or lack therefor) with tests measuring different constructs). Expert-novice comparisons
Response process	"The fit between the construct and the detailed nature of performance... actually engaged in" [8]	Analysis of examinees' or raters' thoughts or actions during assessment (e.g. think-aloud protocol). Assessment security (e.g. prevention of cheating) Quality control (e.g. video capture). Rater training
Consequences	"The impact, beneficial or harmful and intended or unintended, of assessment" [9]	Impact on examinee performance (e.g. downstream effects on board scores, graduation rates, clinical performance, patient safety) Other examinee effects (e.g. test preparation, length of training, stress, anxiety) Definition of pass/fail standard

Table 3 Levels of evidence and levels of recommendation according to European Association of Endoscopic Surgeon guidelines, 2005. Adapted from [4]

	Criteria
Level of evidence	
1a	Systematic reviews (or metaanalyses) containing at least some trials of level 1b evidence, in which results of separate, independently conducted trials are consistent
1b	Randomised controlled trial of good quality and of adequate sample size (power calculation)
2a	Randomised trials of reasonable quality and/or of inadequate sample size
2b	Nonrandomised trials, comparative research (parallel cohort)
2c	Nonrandomised trial, comparative research (historical cohort, literature controls)
3	Nonrandomised, noncomparative trials, descriptive research
4	Expert opinions, including the opinion of Work Group members
Level of recommendation	
1	Based on one systematic review (1a) or at least two independently conducted research projects classified as 1b
2	Based on at least two independently conducted research projects classified as level 2a or 2b, within concordance
3	Based on one independently conducted research project level 2b, or at least two trials of level 3, within concordance
4	Based on one trial at level 3 or multiple expert opinions, including the opinion of Work Group members (e.g. level 4)

preponderance of 'level 4' evidence, it is not possible to yet provide an EAES recommendation supporting the content validity of the LapSim automated training content or assessment tools for gynaecological, general surgery or cardiothoracic surgery.

Internal structure validity

The internal structure validity evidence outlined by included studies is referenced in Table 6. Haidari et al. recorded significant moderate (r 0.42–0.52) correlations in 'mean time' scores across three attempts at a simulated VATS lobectomy, whilst correlations in 'mean blood loss' and 'mean instrument path length' scores were inconsistent [12]. Jensen et al. reported on the use of the newly developed VATSAT tool for assessment of observed simulated VATS performance, and not on the internal structure validity of the built-in LapSim automated metrics [13]. Van Dongen et al. report that LapSim parameter scores correlated with overall assessment scores, though the strength of these correlations are not reported further [45].

Based on evidence reported by included studies, a recommendation (using EAES guidelines) of 2 can be provided with regard to the internal consistency reliability of automated performance metrics in assessing performance in endoscopic surgery.

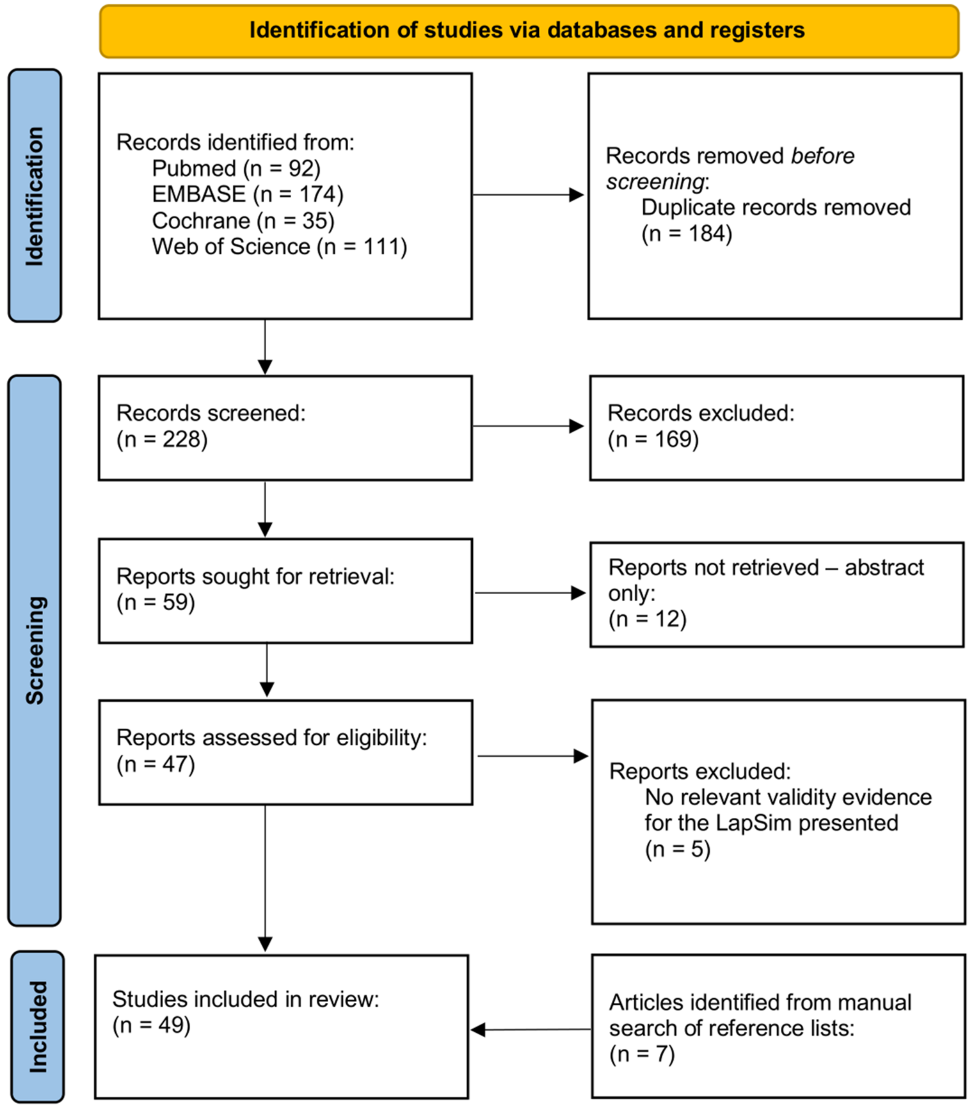

Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart (Adapted from [63]

Relationships with other variables

Case volume

Studies reporting the relationship between LapSim-measured metrics and other relevant performance measures are outline in Table 7. Doneza et al. [11] and Mathews et al. [15] report significant positive associations between at least one objective automated LapSim metrics and self-reported case volume though findings across metrics are inconsistent.

Laparoscopic experience

Fourteen studies were able to demonstrate differences between expert and novice groups across one or more automated LapSim parameters [12, 17, 25, 27, 28, 33, 44, 45, 48, 50, 51, 53, 58, 59]. Sinha et al. [39] and Lin et al. [20] demonstrated the ability of one or more LapSim metrics to discriminate performance amongst participants of differing post-graduate years and training grades respectively.

Several further studies reported significant associations between performance as measured by one or more automated LapSim performance metrics and surgical or laparoscopic experience as measured through self-reported procedural numbers [52] [20] [30] [37] [31] [56] [25, 54, 59]. While no difference in measured baseline metrics between residents and students was noted in Hassan et al.'s study, residents showed improved scores after LapSim training, implying improved skill acquisition as measured by the LapSim [49].

It is important to note that whilst the above studies record score variance related to surgical experience, this finding is inconsistent across variables. No study demonstrated the discriminative ability of all measured LapSim performance metrics, and the relationship between these variables and operative experience differed across studies, populations, tasks and procedures.

Table 4 Characteristics of included studies

Author	Year	Country	Population	Specialty	N
Doneza [11]	2021	USA	Obstetricians and gynaecologists	Gynaecology	108
Haidari [12]	2021	Denmark	Students, trainees, thoracic surgeons	Cardiothoracic	41
Jensen [13]	2018	Denmark	Students, trainees, thoracic surgeons	Cardiothoracic	53
Vapenstad [14]	2017	Norway	Students	NA	30
Mathews [15]	2017	USA	Gynaecologists	Gynaecology	347
Bjerrum [16]	2017	Denmark	Surgeons stratified based on laparoscopic experience	Cardiothoracic	45
Jensen [17]	2016	Denmark	Students, trainees and thoracic surgeons	NA	53
Jensen [18]	2015	Denmark	Thoracic Surgeons performing VATS procedures	Thoracic	103
Alwaal [19]	2015	Canada	Urology residents	Urology	12
Lin [20]	2015	Australia	Students, interns, RMOs, surgical trainees and surgeons	NA	132
Akdemir [21]	2014	Turkey	First and second year residents	Gynaecology	40
Khan [22]	2014	Australia	Students, interns, RMSs, surgical trainees	General	26
Akdemir [23]	2013	Turkey	Postgraduate year-1 gynaecology residents	Gynaecology	22
Xafis [24]	2013	Australia	Students, interns, RMOs, surgical trainees and surgeons	General	228
Lehman [25]	2012	Germany	Attendees at 'Practical Course for Visceral Surgery' with median 3 years professional experience	Not specified	36
Tan [26]	2012	Australia	Students, interns, RMOs, surgical trainees and surgeons	General	228
Shetty [27]	2012	USA	Students, residents, fellows and attendings	General	41
Kovac [28]	2012	Canada	Urology Residents	Urology	15
Palter [29]	2012	Canada	Opinion leaders in colorectal surgery	Colorectal	19
Vapenstad [30]	2012	Norway	Surgeons stratified based on laparoscopic experience	General	47
Panait [31]	2011	USA	Surgical residency applicants	General	51
Calatayud [32]	2010	Denmark	General Surgeons	General	10
van Dongen [33]	2010	Netherlands	Expert surgeons	General	80
Gauger [34]	2010	USA	Surgery Interns	General	14
Tanoue [35]	2010	Japan	Surgical trainees	General	194
Maagaard [36]	2010	Denmark	trainees and senior consultants	Gynaecology	19
Schreuder [37]	2009	Netherlands	Gynaecologists, residents and students	Gynaecology	56
Larsen [38]	2009	Denmark	Gynaecology trainees	Gynaecology	24
Sinha [39]	2008	USA	General Surgery residents	General	33
Hogle [40]	2008	USA	Postgraduate-year-1 residents	General	21
Kundhal [41]	2008	Canada	Surgical residents	General	10
Salgado [42]	2008	USA	Fellowship applicants, minimally invasive surgery	General	8
Panait [43]	2008	USA	Residents and attending surgeons	General	32
Botden [44]	2007	Netherlands	Surgeons stratified based on laparoscopic experience	General	90
Cosman [60]	2007	Australia	Surgical Trainees	General	10
van Dongen [45]	2007	Netherlands	Students, residents, experienced endoscopic surgeons	General	48
Newmark [46]	2007	USA	Medical Students	Gynaecology	47
Hogle [47]	2007	USA	Medical students	NA	29
Ahlberg [48]	2007	Sweden	Surgical residents	General	13
Hassan [49]	2006	Germany	Students and surgical residents	General	24
Woodrum [50]	2006	USA	Students, residents and faculty surgeons	General	34
Larsen [51]	2006	Denmark	Gynaecologists stratified by experience	Gynaecology	32
Aggarwal [52]	2006	UK	Experienced laparoscopic surgeons and novices	General	40
Aggarwal [53]	2006	UK	Gynaecological surgeons	Gynaecology	23
Eriksen [54]	2005	Denmark	Surgeons stratified based on laparoscopic experience	General	24
Hassan [55]	2005	Gernany	Surgeons stratified based on laparoscopic experience	General, Thoracic, Vascular	27
Langelotz [56]	2005	Germany	German surgical society convention attendees	General	115
Sherman [57]	2005	Canada	Students, surgical trainees and faculty surgeons	General	24

Table 4 (continued)

Author	Year	Country	Population	Specialty	N
Duffy [58]	2004	USA	Residents and attending laparoscopic surgeons	General	54

RMO Registered Medical Officer, *NA* not applicable, *NS* not specified

Table 5 Summary of published content validity evidence for the LapSim

Author	Year	N	Population	Content validity evidence
Doneza [11]	2021	108	Obstetricians and gynaecologists	39.8% agreed or strongly agreed that the simulation of three laparoscopic gynaecological tasks was realistic
Haidari [12]	2021	41	Students, trainees, thoracic surgeons	As described by Jensen et al. [18]
Jensen [13]	2018	53	Students, trainees, thoracic surgeons	As described by Jensen et al. [18]
Mathews [15]	2017	347	Gynaecologists	58.1% agreed or strongly agreed that the technique simulations were realistic
Jensen [17]	2016	53	Students, trainees and thoracic surgeons	As described by Jensen et al. [18]
Jensen [18]	2015	103	Thoracic Surgeons performing VATS procedures	The realism of the simulated VATS scenario was rated on a 7-point Likert scale (from 1 – not realistic to 7 – very realistic) by 106 'novice', 'intermediate' and 'experienced participants. The overall realism was rated as a median of '5' by all groups. The Experienced VATS surgeons (N = 26) also rated the realism of the graphical representation of the operating field and the Realism of the movements of the instruments and structures of the lung with median scores of '5'. The similarity of the simulation to a real-life procedure was rated as a '4' [18]. Simulation development and further pilot-testing is also briefly described [18]
Shetty [27]	2012	41	Students, residents, fellows and attendings	70% reported that a laparoscopic camera navigation curriculum using the LapSim was a valid testing tool, with no participants disagreeing or strongly disagreeing
Palter [29]	2012	19	Opinion leaders in colorectal surgery	Delphi process of opinion leaders in colorectal surgery used to generate appropriate tasks for inclusion in a simulation training and assessment curriculum. Consensus was reached for 7 basic tasks and one advanced suturing tasks, allowing for the development of subsequent expert benchmarks
van Dongen [33]	2010	80	Expert surgeons	Delphi consensus. Eight European teams, with experience using the LapSim, defined 8 basic skill exercises and a suturing task to include in a proficiency-based training and assessment curriculum for basic endoscopic surgical skills
Schreuder [37]	2009	56	Gynaecologists, residents and students	Realism of 9 different tasks in gynaecological surgery rated across 11 survey items; the mean score was 3.29 (rated on a 5-point Likert scale), with the realism of the myoma suturing skill (2.43), the tissue reaction on manipulation (2.79) and the appearance of needle and thread (2.87) rated less favourably
Botden [44]	2007	90	Surgeons stratified based on laparoscopic experience	ProMIS (Haptica, Dublin, Ireland) rated as more realistic than the LapSim in both a basic skills task and suturing task. In particular the ProMIS was regarded as having better haptic feedback

Performance on other simulator models

Vapenstad et al. assessed performance in a simulated porcine cholecystectomy by student who reached pre-defined automated assessment criteria on the LapSim and those who did not [14]; the control group achieved significantly better scores than the simulator-trained group [14]. Hogle et al. [40] recorded improved depth perception when performing a

Table 6 Internal consistency validity

Author	Year	Validity Evidence	Key Findings	Level
Haidari [12]	2021	Internal consistency Test–retest reliability	Cronbach's alpha coefficient for standardised items of 0.91. Test–retest reliability: Significant moderate correlations in score parameters across test occasions	2b
Jensen [13]*	2018	Inter-rater reliability Test-re-test reliability	Inter-rater reliability: ICC 0.78 for single measures and 0.91 for average measures*. Test/re-test reliability Pearson's coefficient of 0.70 $p < 0.001$*. G-coefficient 0.79 with two occasions and three raters; the variance of the results was found mainly to be related to the participants and not the raters or assessment occasions.*	2b
Jensen [17]	2016	Internal consistency Test–retest reliability	Cronbach's alpha 0.90—0.91 across 19 items. Significant linear relationship between first and second attempts for 15 of 19 items	2b
Sinha [39]	2008	Internal consistency	On multivariate analysis, correlation of individual parameters to overall task failure were 0.44 for time, 0.49 for path or angle length and 0.88 for tissue damage	2b
van Dongen [45]	2007	Internal consistency	The scores for efficiency, speed and precision (LapSim parameters clustered into three overarching categories) were consistent with the overall score	2b
Hogle [47]	2007	Test–retest reliability	Camera navigation task: ICC 0.35 – 0.72 across parameters, Kappa 0.40 – 0.51. Lifting and grasping task: ICC 0.49 – 0.85 across parameters, Kapps 0.37 – 0.82	3

ICC Intra-Class Coefficient, *G-coefficient* reliability coefficient produced by reliability studies analysed using Generalisability theory

*Jensen et al. do not report on the internal structure validity of automated performance metrics as measured by the LapSim, but rather the use of the VATSAT tool for assessing observed performance on the LapSim

porcine cholecystectomy by residents who had trained using the LapSim, though no significant improvement was noted across other measured domains. In Xafis et al.'s study, participants trained to proficiency on the LapSim in a mobile simulation unit recorded a 23.1% increase in subsequent Fundamentals of Laparoscopic Surgery (FLS) scores compared to baseline data [24], though this finding was not statistically significant (p 0.14) [24]. Similarly, participants in Tan et al.'s study recorded an 11% increase over baseline FLS scores after training to proficiency using the LapSim automated metrics, though analysis beyond descriptive statistics for this finding is not provided [26]. In Newark et al.'s study of medical students, automated performance metrics measured by the LapSim correlated moderately with observed and graded performance on laparoscopic tasks using a box trainer [46].

Pre- vs post- training studies

The relationship between automated LapSim assessment metrics and in-theatre OSATs scores is not further reported. Akdemir et al. report on the ability of the LapSim platform to discriminate between trainees who had undergone a structured laparoscopic training programme using a box trainer and those who had not; the box trainer group showed significantly greater improvements in time and economy of motion scores as assessed by the LapSim compared to their counterparts who had not undergone such training [23].

In summary, there is sufficient evidence to award an EAES recommendation of 2 with respect to the ability of the LapSim to discriminate based on case volume and overall laparoscopic experience, with several studies demonstrating differences in one or more automated metrics across training levels, case volume and overall case numbers. However, it is important to note that the relationships between variables of interest and automated LapSim metrics varies across studies, procedures and tasks. The relationship between LapSim scores and performance using other established and validated simulator models requires further exploration.

Response process validity

Four studies reported specifically on measures to ensure the response process validity of assessments using the LapSim (all prospective observational studies, level 2b) [12, 13, 17, 18]. All studies report validity evidence for the LapSim in the context of assessment of competence in VATS procedures, produced by a research group at the University of Copenhagen. Measures included ensuring that participants had the same level of familiarity with the assessment or VATS module prior to testing, ensuring that instructor was present throughout the entire assessment, and disallowing any hand-on guidance or teaching. Procedures for module or assessment randomisation, rater training and data security are also described [12, 13, 17, 18].

The response process validity of a given assessment is specific to the institutional context and proposed interpretation of the assessment results. It is therefore not possible to provide an EAES recommendation with regard to the LapSim simulator as a standalone device.

Table 7 Construct validity evidence for the LapSim

Author	LapSim assessment tool	Relevant variable(s)	Relevant variable assessment tool	Significant Positive association reported in one or more parameter	Level of evidence
Doneza [11]	Automated	Case volume	Self-reported monthly laparoscopic case volume	Yes	2b
Haidari [12]	Automated	Experience	Self-reported VATS experience	Yes	2b
Jensen [13]	VATSAT	Experience	Self-reported VATS experience	Yes	2b
Vapenstad [14]	Automated	Performance on porcine cholecystectomy model	GOALS	No	2a
Mathews [15]	Automated	Case volume, self-reported skill	Self-reported laparoscopic case volume and competence	Yes	2b
Bjerrum [16]	Automated	Experience	Self-reported appendicectomy experience	Yes	2b
Jensen [17]	Automated	Experience	Self-reported VATS experience	Yes (some)	2b
Jensen [18]	Automated	Experience	Self-reported VATS experience		2b
Alwaal [19]	Automated	Porcine nephrectomy model	Adapted global rating scale	No	3
Lin [20]	Automated	Experience, self-reported competence	Self-reported training level, prior surgical experience, number of procedures performed and assisted, self-reported competence	Yes	2b
Akdemir [23]	Automated	Pre- and post- training	Trainees who underwent structured training using a box-trainer compared to those who did not	Yes	
Xafis [24]	Automated	FLS scores	FLS scores assessed after reaching proficiency on LapSim	No	2b
Lehmann [25]	Automated	Experience	Slef-reported number of laparoscopic cases	Yes	2b
Tan [26]	Automated	FLS scores	FLS scores assessed after reaching proficiency on LapSim	*	2a
Shetty [27]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
Kovac [28]	Automated	Experience	Post-graduate year	Yes	3
Vapenstad [30]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
Panait [31]	Automated	Experience	Training level	Yes	2b
Catalayud [32]	NA	Pre-post training	OSATs	Yes	2a
van Dongen [33]	Automated	Experience	Novices, surgeons in training, surgeons	Yes	2b
Tanoue [35]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
Schreuder [37]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
Sinha [39]	Automated	Experience	Training Year	Yes	2b
Panait [43]	Automated	Experience	Post-graduate year	Yes	2b
Botden [44]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
van Dongen [45]	Automated	Experience	Training level, self-reported laparoscopic experience	Yes	2b
Newmark [46]	Automated	Performance on a laparoscopic box trainer	Time taken, number of 'errors' (pre-defined)	Yes	3
Ahlberg [48]	Automated	Experience	Expert-novice comparisons, defined by experience in advanced laparoscopy	Yes	1b

Table 7 (continued)

Author	LapSim assessment tool	Relevant variable(s)	Relevant variable assessment tool	Significant Positive association reported in one or more parameter	Level of evidence
Hassan [49]	Automated	Experience	Scores compared between residents and medical students	Yes	2b
Woodrum [50]	Automated	Experience	Training level	Yes	2b
Larsen [51]	Automated	Experience	Number of advanced laparoscopic procedures performed (self-reported)	Yes	2b
Aggarwal [52]	Automated	Experience	Self-reported laparoscopic cholecystectomy experience	Yes	1b
Aggarwal [53]	Automated	Experience	Self-reported numbers of laparoscopic procedures performed	Yes	2b
Langelotz [56]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
Eriksen [54]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
Hassan [59]	Automated	Experience	Seld-reported laparoscopic experience	Yes	2b
Sherman [57]	Automated	Experience	Self-reported laparoscopic experience	Yes	2b
Duffy [58]	Automated	Experience	Training level	Yes	2b

GOALS Global Operative Assessment of Laparoscopic Tools, *FLS* Fundamentals of Laparoscopic Surgery, *VATS* Video-Assisted Thoracoscopic Surgery, *NA* Not applicable

*Descriptive statistics only provided

Consequential validity

Eight studies report consequential validity evidence relevant to the use of the LapSim in assessment of skill or competence [12, 13, 17, 33, 38, 42, 48, 52]. Four studies report procedures for generating pass/fail benchmarks, summarised in Table 8.

Larsen et al. (level of evidence 1b) demonstrated that participants who had trained to proficiency benchmarks on the LapSim demonstrated improved performance as assessed in the operating theatre during a laparoscopic salpingectomy, compared to trainees who did not undergo such training or assessment on the LapSim ($p < 0.001$) [38]. In Ahlberg et al.'s randomised controlled trial (level of evidence 1b), residents who reached expert benchmarks on LapSim metrics made significantly fewer errors in the operating room during their first ten laparoscopic cholecystectomy procedures than controls did ($p = 0.0037$) [48]. Residents passing these bench-marks also operated 58% faster than their colleagues undergoing training and assessment without the LapSim [48]. These findings are supported by a further study by Cosman et al. (level of evidence 1b) [60]; trainees who trained using pre-defined LapSim proficiency targets made significantly fewer errors in subsequent laparoscopic cholecystectomy performance than those who did not train using the LapSim [60]. Conversely, Hogle et al [40] could not demonstrate any association between time taken to complete basic skills training on the LapSim and subsequent in-theatre

Table 8 Reported methods of determining pass/fail benchmarks

Author	Procedure or task	Method of pass/fail score generation	False positive	False negative
Haidiri [12]	VATS lobectomy	Contrasting groups' method	1/22 (4.54%)	6/9 (66.67%)
Jensen [13]	VATS lobectomy	Contrasting groups' method (VATSAT tool)	5/17 (29.41%)	6/14 (42.86%)
Jensen [17]	VATS lobectomy	Mean expert score plus one standard deviation	0/17 (0.00%)	4/14 (29%)
Van Dongen [17]	Basic laparoscopic skills	Mean expert score plus two standard deviations	NR	NR

False positive the proportion of novice participants who would pass the assessment with the given pass/fail benchmarks, *False negative* the number of expert participants who would fail the assessment with the given pass/fail benchmarks, *VATSAT* Video-Assisted Thoracoscopic Surgery lobectomy Assessment Tool), *NR* not recorded laparoscopic cholecystectomy performance; clinical performance data was only available for six of the 21 recruited residents (level of evidence 3). Salgado et al. explore the use of technical skill assessment using the LapSim in the selection process for surgical trainees wishing to pursue a fellowship in minimally invasive surgery [42] (level of evidence 3). One candidate (of eight) did not achieve proficiency criteria on a pre-defined ‘easy’ level, one on ‘medium’ level and three with benchmarks set at a ‘difficult’ level [42]. There was no correlation between scores achieved on the LapSim and scores awarded in an interview assessing each candidate’s curriculum vitae, published research work and laparoscopic or endoscopic experience [42].

A further study by Gauger et al. (level of evidence 1b) supports the use of proficiency targets in LapSim training; interns randomised to undergo training with pre-defined proficiency targets performed better in subsequent ‘live’ laparoscopic cholecystectomy assessments compared to those who did not [34]. It is important to note that participants in this study were not required to reach these proficiency targets prior to final assessment, though trainees in the intervention group were significantly more likely to reach these targets than those in the control group [34]. Akdemir et al. report superior performance in live laparoscopic salpingectomy by trainees trained using either the LapSim or a box trainer, for one hour a week over five weeks, compared to controls [21]. No difference in performance was recorded between the LapSim- or Box trainer- trained cohorts, though it is important to note that neither group trained to pre-defined proficiency benchmarks [21]. In a randomised controlled trial by Catalayud et al. [34], surgeons who underwent rapid (15-min) ‘warm-up’ training prior to performance of a ‘live’ laparoscopic cholecystectomy performed better significantly better as measured by OSATs grading of videotaped performance (mean total OSATs score (28.5 vs 19.25, $p=0.042$) compared to their counterparts who did not partake in any warm-up training [34].

Three good quality randomised controlled trials have demonstrated that performance assessment metrics on the LapSim correlate with improved performance in the operating room (EAES level of recommendation 1) [38, 48, 60]. Two further randomised controlled trials have further demonstrated the benefit of LapSim training, without requirements for reaching proficiency targets, on subsequent live operative performance [32, 34]. Akdemir et al.’s study however, reports similar improvements in performance using a simple box trainer compared to the LapSim [21].

Discussion

This review outlines and evaluated the current validity evidence for the LapSim laparoscopic simulator, particularly with regard to its intended use as an assessment tool and method. Processes for generating content validity varied across included studies [11–13, 15, 17, 18, 27, 29, 33, 37, 44]. In particular, two studies used expert consensus methodology to define tasks and procedures for inclusion in a colorectal surgery and basic surgical skills curriculum respectively [29, 33]. Data from these studies can inform the development and implementation of future LapSim-based assessment curricula.

Perceptions regarding the fidelity of the LapSim varied across studies. Whilst participants in a recent study evaluating the use of the LapSim platform in assessing VATS performance rated the overall realism of the platform favourably [18], another recent study recorded that only 39.8% of participants agreed or strongly agreed that the LapSim reproductions of gynaecological tasks were realistic [11]. Several studies from a Copenhagen University research group report a rigorous process of content validity generation and software development for the LapSim VATS module [12, 13, 17, 18], in contrast with a relative lack of similar studies in relating to other endoscopic surgical disciplines such as gynaecology and general surgery. The perceived realism of simulations likely varies across LapSim procedures and models. The acceptability of the LapSim platform as a valid testing tool is only explored by one study, though results are favourable [27].

Internal consistency data was not consistently reported across studies. In particular, no studies used test-item statistics or factor analysis to evaluate the contribution of individual automated metrics to the overall construct or pass mark. The relationship of LapSim assessment scores to other relevant variables measuring a similar construct was by far the most commonly presented validity evidence (Table 6). All studies examining the relationship between laparoscopic surgical experience or case volume and automated LapSim performance metrics were able to demonstrate the discriminative ability of one or more parameter [11, 12, 15, 17, 20, 27, 28, 30, 31, 33, 35, 37, 39, 44, 45, 48–54, 56, 58]. However, it is important to note that the discriminative ability of automated metrics such as path length or tissue damage varied widely and performed inconsistently across included studies. A systematic review and meta-analysis of expert-novice comparisons as assessed by these automated metrics is beyond the scope of this review, and would be challenging given the heterogeneity of studied populations, procedures and tasks.

Interestingly, several studies were unable to identify a positive relationship between performance as assessed by the LapSim, and simulated performance assessed using other simulators or animal models [14, 19, 24, 26]. Vapenstad et al. assessed candidate performance on a porcine model using a modified GOALS (Global Operative Assessment of Laparoscopic Skills) tool [14]. Alwaal et al. 2015 also utilised porcine models and a pre-defined rating scale that again appears to be a modified GOALS tool with additional domains [19]. These studies present negative evidence regarding the relationship between LapSim assessment metrics and other measured variables, though the increased variance introduced by differing anatomy in animal models and the use of newly modified assessment tools may affect the overall validity of this evidence. The use of validated assessment tools to evaluate performance on both the LapSim and other simulated or animal models should help to clarify the transferability of skills across platforms in future studies. Conversely, three randomised controlled trials have demonstrated that trainees who reach pre-defined expert-level benchmarks on the LapSim demonstrate improved performance in the operating theatre on human patients [38, 48, 60]. Such evidence supports the development of future studies exploring the use of the LapSim in high-stakes assessment scenarios such as trainee progression, certification and autonomy granting. Given the preponderance of evidence supporting proficiency-based training in surgical education [61], it is important that future studies explore the institutional, financial and other barriers that have thus far prevented simulation-based training and assessment criteria, regardless of simulator platform, to become the norm in surgical training [62].

A number of important consequential validity evidence domains remain unexplored. In particular, the cost-effectiveness of utilising the LapSim as an assessment tool is not examined by identified studies. A single LapSim platform costs approximately $55,000 (including set-up) [19]. Further costs are likely incurred with respect to platform maintenance and updates, whilst variable costs will be incurred related to the number of trainees to be assessed, associated time and staffing resource requirements, and administrative burden [19]. The potential benefits are diverse and likely difficult to comprehensively evaluate. For example, training programmes might hypothesise that ensuring a benchmarked performance standard using LapSim assessments could lead to improved patient outcomes and associated cost-benefits, less time resources spent on in-theatre assessment and potentially reduced training-time in line with the ideals of competency-based education. Whilst weighing the relative costs and benefits of utilising the LapSim is therefore challenging, it is essential that training programmes have at least an expectation of associated costs so that they can make purchasing decisions regarding the appropriate training or assessment platform for the intended use. Furthermore, the impact on trainees of utilising the LapSim (or any other virtual reality platform) as an assessment tool is an important consideration. The introduction of any new assessment tool or method requires an analysis of the time spent by trainees to prepare for such assessments, the stress or anxiety caused by assessments and the associated costs (both financial and psychological) of failure and remediation.

European Association of Endoscopic Surgeon (EAES) guidelines were used to evaluate the current evidence landscape for the LapSim. Based on the available evidence, an EAES recommendation of 2 can be provided regarding the relationship between automated performance metrics and other relevant variables measuring a similar construct, namely laparoscopic experience and case volume. A recommendation level of 1 (the highest recommendation) can be given regarding the ability of the LapSim automated metrics to predict future performance in the operating theatre (consequential validity). However, the above review highlights the limits of these guidelines. In particular, the EAES guidelines were published before the widespread adaptation of modern validity frameworks (e.g. Messick's or Kane's framework). Whilst 'expert opinion' is awarded a low level of evidence grading, such methodologies are well established in the context of content validity generation. Similarly, randomised controlled trials may not be the most appropriate research method for generating much of the validity evidence required for the use of virtual reality platforms in assessment. The authors call for updated, standardised guidelines based on a modern validity evidence framework for evaluating the available evidence pertaining to individual simulator models, so that decisions on procurement by educational bodies can be made, and areas for future research more easily identified.

Furthermore, there is a need for standardised guidelines regarding the reporting of virtual reality based simulation studies. These should include a definition of the term 'expert' for the purpose of benchmarking, along with a list of recommended reporting items pertaining to Messick's validity framework or a similar contemporary approach to validity evidence collation.

Limitations

A systematic review and meta-analysis of included outcomes is precluded by the design of this paper as a scoping review, and the significant heterogeneity of included studies with respect to populations, interventions, comparator groups and outcomes. As such, a critical analysis of all automated performance parameters across tasks and procedures as used by the LapSim was not performed. This article evaluates the evidence pertaining to one specific simulator model only, and does not aim to make comparisons with other virtual reality laparoscopic simulators. The LapSim has gone through several design updates and software iterations since its inception; in particular, later versions have added haptic feedback. This could conceivably affect the validity of its use as both a training and assessment platform. It is possible that findings from earlier studies, in particular with regard to findings related to haptics and realism, may no longer be relevant.

Conclusion

Available evidence supports the positive, though inconsistent, relationship between automated performance assessment metrics as measured by the LapSim and relevant proxy measures of operative skill; namely laparoscopic experience and case volume. Whilst studies have demonstrated the relationship between the achievement of expert-level benchmarks on the LapSim platform and future in-theatre performance, no study has yet examined the relationship between the achievement of these benchmarks and subsequent patient outcome. This remains an area for future study development. Further studies are required to investigate the impact of implementing simulation-based training and assessment programmes in post-graduate surgical training, at a programme- or national-level. The authors call for updated guidelines with regard to the evaluation of validity evidence available for virtual reality models with reference to modern validity frameworks, along with standardised reporting guidelines for virtual reality simulator studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00464-022-09593-0.

Funding Financial support was provided as part of a higher degree (C Toale) by the Royal College of Surgeons/Hermitage Medical Clinic Strategic Academic Recruitment (StAR MD) programme.

Declarations

Disclosures Mr Conor Toale, Dr Marie Morris and Mr Dara O. Kavanagh have no conflicts of interest or financial ties to disclose.

References

1. Hong M, Rozenblit JW, Hamilton AJ (2021) Simulation-based surgical training systems in laparoscopic surgery: a current review. Virtual Real 25:491–510
2. Toale C, Morris M, Kavanagh D (2021) Perceptions and experiences of simulation-based assessment of technical skill in surgery: a scoping review. Am J Surg 222:723–730
3. McDougall EM (2007) Validation of surgical simulators. J Endourol 21:244–247
4. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ (2005) Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 19:1523–1532
5. Fairhurst K, Strickland A, Maddern G (2011) The LapSim virtual reality simulator: promising but not yet proven. Surg Endosc 25:343–355
6. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MD, Horsley T, Weeks L (2018) PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 169:467–473
7. Smyer P (2011) The medical education research study quality instrument (MERSQI)
8. Hambleton RK, Arrasmith G, Sheehan DS, Grobe RP, Hathaway W, Doherty V, Forbes D, Houser R, Ingebo G, Kingsbury G (1986) Standards for educational and psychological testing: six reviews. JSTOR
9. Cook DA, Lineberry M (2016) Consequences validity evidence: evaluating the impact of educational assessments. Acad Med 91:785–795
10. Cook DA, Hatala R (2016) Validation of educational assessments: a primer for simulation and beyond. Adv Simul 1:31
11. Doneza JA, Palvia V, Lerner VT, Overbey J, Levie M, Brodman M, Ascher-Walsh C (2021) Correlation of surgical case volume and fellowship training with performance on simulated procedural tasks. Am J Obstet Gynecol 225:548.e541-548.e510
12. Haidari TA, Bjerrum F, Hansen HJ, Konge L, Petersen RH (2022) Simulation-based VATS resection of the five lung lobes: a technical skills test. Surg Endosc 36:1234–1242
13. Jensen K, Hansen HJ, Petersen RH, Neckelmann K, Vad H, Møller LB, Pedersen JH, Konge L (2019) Evaluating competency in video-assisted thoracoscopic surgery (VATS) lobectomy performance using a novel assessment tool and virtual reality simulation. Surg Endosc 33:1465–1473
14. Våpenstad C, Hofstad EF, Bø LE, Kuhry E, Johnsen G, Mårvik R, Langø T, Hernes TN (2017) Lack of transfer of skills after virtual reality simulator training with haptic feedback. Minim Invasive Ther Allied Technol 26:346–354
15. Mathews S, Brodman M, D'Angelo D, Chudnoff S, McGovern P, Kolev T, Bensinger G, Mudiraj S, Nemes A, Feldman D, Kischak P, Ascher-Walsh C (2017) Predictors of laparoscopic simulation performance among practicing obstetrician gynecologists. Am J Obstet Gynecol 217:596.e591-596.e597
16. Bjerrum F, Strandbygaard J, Rosthøj S, Grantcharov T, Ottesen B, Sorensen JL (2017) Evaluation of procedural simulation as a training and assessment tool in general surgery-simulating a laparoscopic appendectomy. J Surg Educ 74:243–250
17. Jensen K, Bjerrum F, Hansen HJ, Petersen RH, Pedersen JH, Konge L (2017) Using virtual reality simulation to assess competence in video-assisted thoracoscopic surgery (VATS) lobectomy. Surg Endosc 31:2520–2528
18. Jensen K, Bjerrum F, Hansen HJ, Petersen RH, Pedersen JH, Konge L (2015) A new possibility in thoracoscopic virtual reality simulation training: development and testing of a novel virtual reality simulator for video-assisted thoracoscopic surgery lobectomy. Interact Cardiovasc Thorac Surg 21:420–426
19. Alwaal A, Al-Qaoud TM, Haddad RL, Alzahrani TM, Delisle J, Anidjar M (2015) Transfer of skills on LapSim virtual reality laparoscopic simulator into the operating room in urology. Urol Ann 7:172–176
20. Lin D, Pena G, Field J, Altree M, Marlow N, Babidge W, Hewett P, Maddern G (2016) What are the demographic predictors in laparoscopic simulator performance? ANZ J Surg 86:983–989
21. Akdemir A, Sendağ F, Oztekin MK (2014) Laparoscopic virtual reality simulator and box trainer in gynecology. Int J Gynaecol Obstet 125:181–185
22. Khan MW, Lin D, Marlow N, Altree M, Babidge W, Field J, Hewett P, Maddern G (2014) Laparoscopic skills maintenance: a randomized trial of virtual reality and box trainer simulators. J Surg Educ 71:79–84

23. Akdemir A, Ergenoğlu AM, Yeniel A, Sendağ F (2013) Conventional box model training improves laparoscopic skills during salpingectomy on LapSim: a randomized trial. J Turk Ger Gynecol Assoc 14:157–162
24. Xafis V, Babidge W, Field J, Altree M, Marlow N, Maddern G (2013) The efficacy of laparoscopic skills training in a Mobile Simulation Unit compared with a fixed site: a comparative study. Surg Endosc 27:2606–2612
25. Lehmann KS, Gröne J, Lauscher JC, Ritz JP, Holmer C, Pohlen U, Buhr HJ (2012) Simulation training in surgical education—application of virtual reality laparoscopic simulators in a surgical skills course. Zentralbl Chir 137:130–137
26. Tan SC, Marlow N, Field J, Altree M, Babidge W, Hewett P, Maddern GJ (2012) A randomized crossover trial examining low- versus high-fidelity simulation in basic laparoscopic skills training. Surg Endosc 26:3207–3214
27. Shetty S, Panait L, Baranoski J, Dudrick SJ, Bell RL, Roberts KE, Duffy AJ (2012) Construct and face validity of a virtual reality-based camera navigation curriculum. J Surg Res 177:191–195
28. Kovac E, Azhar RA, Quirouet A, Delisle J, Anidjar M (2012) Construct validity of the LapSim virtual reality laparoscopic simulator within a urology residency program. Can Urol Assoc J 6:253–259
29. Palter VN, Graafland M, Schijven MP, Grantcharov TP (2012) Designing a proficiency-based, content validated virtual reality curriculum for laparoscopic colorectal surgery: a Delphi approach. Surgery 151:391–397
30. Våpenstad C, Hofstad E, Bø LE, Chmarra M, Kuhry E, Johnsen G, Mårvik R, Langø T (2012) Limitations of haptic feedback devices on construct validity of the LapSim((R)) virtual reality simulator. Surg Endosc 27
31. Panait L, Larios JM, Brenes RA, Fancher TT, Ajemian MS, Dudrick SJ, Sanchez JA (2011) Surgical skills assessment of applicants to general surgery residency. J Surg Res 170:189–194
32. Calatayud D, Arora S, Aggarwal R, Kruglikova I, Schulze S, Funch-Jensen P, Grantcharov T (2010) Warm-up in a virtual reality environment improves performance in the operating room. Ann Surg 251:1181–1185
33. van Dongen KW, Ahlberg G, Bonavina L, Carter FJ, Grantcharov TP, Hyltander A, Schijven MP, Stefani A, van der Zee DC, Broeders IA (2011) European consensus on a competency-based virtual reality training program for basic endoscopic surgical psychomotor skills. Surg Endosc 25:166–171
34. Gauger PG, Hauge LS, Andreatta PB, Hamstra SJ, Hillard ML, Arble EP, Kasten SJ, Mullan PB, Cederna PS, Minter RM (2010) Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. Am J Surg 199:72–80
35. Tanoue K, Uemura M, Kenmotsu H, Ieiri S, Konishi K, Ohuchida K, Onimaru M, Nagao Y, Kumashiro R, Tomikawa M, Hashizume M (2010) Skills assessment using a virtual reality simulator, LapSim, after training to develop fundamental skills for endoscopic surgery. Minim Invasive Ther Allied Technol 19:24–29
36. Maagaard M, Sorensen JL, Oestergaard J, Dalsgaard T, Grantcharov TP, Ottesen BS, Larsen CR (2011) Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. Surg Endosc 25:722–727
37. Schreuder HW, van Dongen KW, Roeleveld SJ, Schijven MP, Broeders IA (2009) Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. Am J Obstet Gynecol 200:540.e541-548
38. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, Schroeder TV, Ottesen BS (2009) Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ 338:b1802
39. Sinha P, Hogle NJ, Fowler DL (2008) Do the laparoscopic skills of trainees deteriorate over time? Surg Endosc 22:2018–2025
40. Hogle NJ, Widmann WD, Ude AO, Hardy MA, Fowler DL (2008) Does training novices to criteria and does rapid acquisition of skills on laparoscopic simulators have predictive validity or are we just playing video games? J Surg Educ 65:431–435
41. Kundhal PS, Grantcharov TP (2009) Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. Surg Endosc 23:645–649
42. Salgado J, Grantcharov TP, Papasavas PK, Gagne DJ, Caushaj PF (2009) Technical skills assessment as part of the selection process for a fellowship in minimally invasive surgery. Surg Endosc 23:641–644
43. Panait L, Bell RL, Roberts KE, Duffy AJ (2008) Designing and validating a customized virtual reality-based laparoscopic skills curriculum. J Surg Educ 65:413–417
44. Botden SM, Buzink SN, Schijven MP, Jakimowicz JJ (2007) Augmented versus virtual reality laparoscopic simulation: what is the difference? A comparison of the ProMIS augmented reality laparoscopic simulator versus LapSim virtual reality laparoscopic simulator. World J Surg 31:764–772
45. van Dongen KW, Tournoij E, van der Zee DC, Schijven MP, Broeders IA (2007) Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 21:1413–1417
46. Newmark J, Dandolu V, Milner R, Grewal H, Harbison S, Hernandez E (2007) Correlating virtual reality and box trainer tasks in the assessment of laparoscopic surgical skills. Am J Obstet Gynecol 197:546.e541–544
47. Hogle NJ, Briggs WM, Fowler DL (2007) Documenting a learning curve and test-retest reliability of two tasks on a virtual reality training simulator in laparoscopic surgery. J Surg Educ 64:424–430
48. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA 3rd, Ramel S, Smith CD, Arvidsson D (2007) Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg 193:797–804
49. Hassan I, Gerdes B, Koller M, Langer P, Rothmund M, Zielke A (2006) Clinical background is required for optimum performance with a VR laparoscopy simulator. Comput Aided Surg 11:103–106
50. Woodrum DT, Andreatta PB, Yellamanchilli RK, Feryus L, Gauger PG, Minter RM (2006) Construct validity of the LapSim laparoscopic surgical simulator. Am J Surg 191:28–32
51. Larsen CR, Grantcharov T, Aggarwal R, Tully A, Sørensen JL, Dalsgaard T, Ottesen B (2006) Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc 20:1460–1466
52. Aggarwal R, Grantcharov TP, Eriksen JR, Blirup D, Kristiansen VB, Funch-Jensen P, Darzi A (2006) An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg 244:310–314
53. Aggarwal R, Tully A, Grantcharov T, Larsen CR, Miskry T, Farthing A, Darzi A (2006) Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy. BJOG 113:1382–1387
54. Eriksen JR, Grantcharov T (2005) Objective assessment of laparoscopic skills using a virtual reality stimulator. Surg Endosc 19:1216–1219
55. Hassan I, Sitter H, Schlosser K, Zielke A, Rothmund M, Gerdes B (2005) A virtual reality simulator for objective assessment of surgeons' laparoscopic skill. Chirurg 76:151–156
56. Langelotz C, Kilian M, Paul C, Schwenk W (2005) LapSim virtual reality laparoscopic simulator reflects clinical experience in German surgeons. Langenbecks Arch Surg 390:534–537
57. Sherman V, Feldman LS, Stanbridge D, Kazmi R, Fried GM (2005) Assessing the learning curve for the acquisition of

laparoscopic skills on a virtual reality simulator. Surg Endosc 19:678–682
58. Duffy AJ, Hogle NJ, McCarthy H, Lew JI, Egan A, Christos P, Fowler DL (2005) Construct validity for the LAPSIM laparoscopic surgical simulator. Surg Endosc 19:401–405
59. Hassan I, Sitter H, Schlosser K, Zielke A, Rothmund M, Gerdes B (2005) Laparoskopiesimulator. Chirurg 76:151–156
60. Cosman PH, Hugh TJ, Shearer CJ, Merrett ND, Biankin AV, Cartmill JA (2007) Skills acquired on virtual reality laparoscopic simulators transfer into the operating room in a blinded, randomised, controlled trial. Stud Health Technol Inform 125:76–81
61. Mazzone E, Puliatti S, Amato M, Bunting B, Rocco B, Montorsi F, Mottrie A, Gallagher AG (2021) A systematic review and meta-analysis on the impact of proficiency-based progression simulation training on performance outcomes. Ann Surg 274:281–289
62. Zevin B, Aggarwal R, Grantcharov TP (2014) Surgical simulation in 2013: why is it still not the standard in surgical training? J Am Coll Surg 218:294–301
63. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://doi.org/10.1136/bmj.n71