The American Journal of Surgery (2010) 199, 529–541

The American Journal of Surgery®

Review

Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees?

Anthony S. Thijssen, M.D.[a], Marlies P. Schijven, M.D., Ph.D., M.H.Sc.[b,*]

[a]Department of Surgery, University Medical Center Utrecht, Utrecht, The Netherlands; [b]Department of Surgery, Academic Medical Center Amsterdam, Amsterdam, The Netherlands

KEYWORDS:
Virtual reality;
Laparoscopy;
Skills assessment;
Education

Abstract

BACKGROUND: A demand for safe, efficient laparoscopic training tools has prompted the introduction of virtual reality (VR) laparoscopic simulators, which might be used for performance assessment. The purpose of this review is to determine the value of VR metrics in laparoscopic skills assessment.

DATA SOURCES: An exhaustive search of the MEDLINE and EMBASE databases was performed to identify publications concerning construct, concurrent and predictive validation of VR simulators. Of 643 publications found, 42 were included in this review. Studies into all 3 types of validation showed a large heterogeneity in study design. Although concurrence of VR metrics with box trainer metrics, mental aptitude tests, and in vivo surgical performance was generally weak, several metrics demonstrated construct validity in selected simulators.

CONCLUSIONS: Using the right simulator, tasks, and metrics, trainees' and experts' laparoscopic skills can reliably be compared. However, VR simulators cannot yet predict levels of real life surgical skills.

Since the introduction of laparoscopic surgery, it has become clear that different skills are needed to successfully complete a laparoscopic procedure than those required for "open" surgery. Patient safety concerns have generated the need for efficient training instruments that can prepare future laparoscopic surgeons for the operating room (OR) without the risk of harming patients.[1] In recent years, several computer-based virtual reality (VR) laparoscopy simulators have been developed to accommodate this need. Typically, VR simulators measure and store several parameters during performance of a simulator task. These parameters or "metrics" form the basis for the assessment of the trainee's performance.

If assessment of competence is to be carried out using a simulator, these metrics must be reliable and valid. The past few years have seen many publications concerning validation of VR laparoscopy simulators. Within this body of literature 5 types of validity can be distinguished: face validity, content validity, construct validity, concurrent validity, and predictive validity.[2] Face and content validity are subjective qualities of a simulator, whereas construct, concurrent, and predictive validity provide quantitative measures of validity for the metrics employed by the simulator.[3] Construct validity is achieved when a statistically significant difference in performance is measured between groups with different levels of prior laparoscopic experience.[4] Concurrent validity is achieved when there is a strong correla-

* Corresponding author. Tel.: +31(0) 205662166; fax: +31(0) 206914858.
E-mail address: m.p.schijven@amc.uva.nl

Manuscript received July 7, 2008; revised manuscript April 14, 2009

0002-9610/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.amjsurg.2009.04.015

tion between performance on a VR simulator and on an established form of laparoscopic assessment, such as a box trainer.[3] Simulator metrics display predictive validity when they show a strong correlation with objective assessment of in vivo surgical skill.[5]

The aims of this review are to present an overview of the metrics and scoring systems used by contemporary VR laparoscopy simulators in trainee assessment and to identify and assess the evidence concerning their validity.

Materials and Methods

Overview of metrics employed by VR simulators

VR laparoscopy simulators which were commercially available or in widespread use for training and research purposes were included in this review. The selected systems were Delltatech Simendo, Haptica ProMIS, Mentice Procedicus MIST, RealSim Systems LTS2000/ISM60 and LTS3E, Select-IT VSOne, Simbionix LAP Mentor, SimSurgery SEP, Surgical Science LapSim, Verefi Technologies EndoTower, RapidFire/SmartTutor, and Head2Head and Xitact LS 500. To identify which metrics were featured by these simulators, the manufacturers were contacted via e-mail with a request to provide documentation concerning the performance assessment systems. Furthermore, the manufacturers' websites were searched for documentation on the simulator metrics if there was no response to the e-mail request. An overview was compiled incorporating information concerning the design of the apparatus, the type of skills being trained and assessed, the metrics being recorded, and details of any featured scoring systems.

Systematic review of validation studies

The MEDLINE and EMBASE literature databases were searched for publications concerning validation of VR metrics and composite scores. The search syntaxes for both databases were designed to be highly sensitive rather than specific: they consisted of the names of the simulators and their manufacturers, as well as synonyms for "virtual reality" and "laparoscopy." The MEDLINE database was searched without limits concerning publication dates. In EMBASE, the search options "All Years" and "EMBASE only" were selected. The titles and abstracts of the publications identified by the search were screened for relevance. For potentially relevant articles the full text was obtained and read. The articles' literature references were checked for any publications of interest that were not encountered during the search.

A selection of publications was made based on the following inclusion criteria: (1) studies into either construct, concurrent, or predictive validation of metrics/scores in VR laparoscopy simulators; (2) articles presenting original data; and (3) English or Dutch language. The exclusion criteria were (1) studies into nonlaparoscopic simulators (eg, flexible endoscopic, arthroscopic, ureteroscopic, robotic surgery); (2) reviews, meta-analyses, surveys, and opinion articles; (3) insufficient description of study design; (4) insufficient description of study group characteristics (ie, number of participants, level of laparoscopic experience); (5) incomplete results reported; and (6) incomplete or inadequate statistical analysis (eg, no P values or confidence intervals reported for differences in performance). The same inclusion and exclusion criteria were applied to validation studies that were encountered during the process of identifying the assessment metrics.

The following data were extracted from the included studies concerning construct validity: the first author's name, the year of publication, the name of the simulator under study, the number of participants and their level of training, the number of laparoscopic procedures previously performed by the participants, the tasks used for measuring VR performance, the metrics/scores that were being validated, and the statistical differences in performance between groups of participants.

Construct validity was rated according to the following criteria: statistically significant differences ($P < .05$) in outcome between groups of different laparoscopic experience (eg, experts vs novices) in all reported tasks were awarded "full construct validity." Statistically significant differences between some but not all groups or for some but not all reported tasks were awarded "partial construct validity." In case of no statistically significant differences between groups "no construct validity" was awarded.

The data extracted from studies into concurrent validity and predictive validity were the first author's name, the year of publication, the name of the VR simulator under study, the number of participants and their level of training, the number of laparoscopic procedures previously performed by the participants, the tasks used for measuring VR performance, the tasks used for measuring box trainer/aptitude test performance (concurrent validity) and OR performance (predictive validity), the metrics that were being validated, and the statistical correlations between VR performance and box trainer/aptitude test or in vivo laparoscopic performance.

Results

Table 1 provides an overview of the simulators under study and the metrics they employ. The process of searching the literature for validation studies and selecting the publications for review is outlined in Figure 1. The MEDLINE search yielded 578 results, and the EMBASE search, 405. After elimination of duplicates, 643 unique publications remained. Screening the titles and abstracts revealed 91 potentially relevant articles. The reasons for excluding the remaining 552 publications are described in Figure 1.

All 91 full-text articles were read closely and the inclusion and exclusion criteria were applied; 40 publications

Table 1 Overview of the metrics featured by the simulators under study

	Parameters						
Simulator	Time	Errors	Path length	Score	Weighting of metrics	Adjustible weighting	Additional metrics reported in validation studies
Simendo	v	v	v	+	+	+	—
ProMIS	v	+	v	+	?	?	Smoothness,[38] number of movements[10]
MIST-VR/Procedicus	v	v	+	v	+	+	Economy of movements,[16,18–20,39] economy of diathermy[16,18,19,39]
LTS2000/ISM60	+	+	−	+	?	?	—
LTS3E	+	+	−	+	?	?	—
VSOne/VEST	+	+	+	?	?	?	—
LAP Mentor	v	+	+	v	?	?	Economy of movements, number of movements, speed[22]
SEP	+	+	+	+	+	+	—
LapSim Basic Skills	+	+	+	+	+	+	Angular path[24]
LapSim Dissection	+	+	+	+	+	+	Angular path, dissected volume[28]
EndoTower	+	+	+	+	?	?	—
RapidFire/SmartTutor	+	+	?	+	?	?	—
Head2Head	+	+	?	+	?	?	—
LS500	+	+	+	+	+	−	—

v = metric is featured by simulator and has been subjected to validation testing of which results have been published (regardless of outcome); + = featured by simulator, not subjected to validation testing; − = not featured by simulator; ? = unknown: information not provided by manufacturer, not reported in literature.

were excluded because no validation of performance metrics was reported. Of the remaining 51 publications, 40 concerned construct validity of VR metrics.[4,6–45] Of these, 6 papers were excluded for various reasons described in Figure 1.[13,14,25,41,44,45] Furthermore, 2 studies into concurrent validity[9,46] and 1 concerning predictive validity[25] were excluded.

The final selection comprised a total of 42 publications included in this review: 34 studies concerning construct validity, 5 concerning concurrent validity, and 5 concerning predictive validity. One publication reported construct as well as concurrent validity[6] and 1 paper concerned predictive as well as construct validity.[7] Checking the reference lists of the full-text publications did not yield any additional papers of interest. The results of the data extraction and analysis are presented below.

Construct validity

An overview of the data extracted from all 34 construct validity studies is shown in Table 2. The Surgical Science LapSim was studied most extensively (12 studies), followed by Mentice MIST-VR (8 studies) and Haptica ProMIS (4 studies). For Mentice Procedicus MIST, Realsim Systems LTS2000/ISM60, Realsim Systems LTS 3E, Select-IT VSOne/VEST, Verifi Rapidfire/SmartTutor, and Verifi Head2Head, no studies into construct validity could be identified.

Numbers of participants ranged from 8[20] to 307.[37] The participants' laparoscopic experience ranged from more than 1,000 procedures performed[12] to no experience at all. In most publications, participants were categorized as "novices," "intermediates," or "experts" based on their laparoscopic experience, although other denominators were used as well. The level of laparoscopic experience was in most cases defined by the number of procedures a subject had performed before participating in the study. Some studies based the level of experience on the number of laparoscopic procedures per year, while others reported the total number of procedures during the subjects' professional career. The number of procedures required to be considered an expert ranged from over 50 (career total) to over 100 per year. Laparoscopic novices' experience ranged from no laparoscopic experience at all to anything under 10 procedures. Some papers specified the types of procedures that were counted to categorize experience, while most did not. Other criteria for expertise were attending a surgical residency program,[23,24,31] the number of years of surgical residency training,[7,10,15,35,39] attending advanced laparoscopic fellowship training,[11,15,39] functioning as an attending surgeon[11,31,35,39,47] or participating in either a basic or an advanced skills course.[37]

Study design characteristics that also differed substantially between studies were the selections of tasks under study (different tasks in different simulators but also different tasks studied in the same type of simulator) and the number of times tasks were repeated.

Most simulator tasks used to investigate construct validity were basic skills tasks: procedural tasks were employed by LapSim[27–29] and Xitact LS 500.[4,37] The metric most frequently studied was time to completion of a task (26 studies), followed by various types of errors (18 studies) and instrument path length (17 studies). Evaluation of validity of composite scores based on more than 1 metric was reported in 14 publications. In several publications concern-

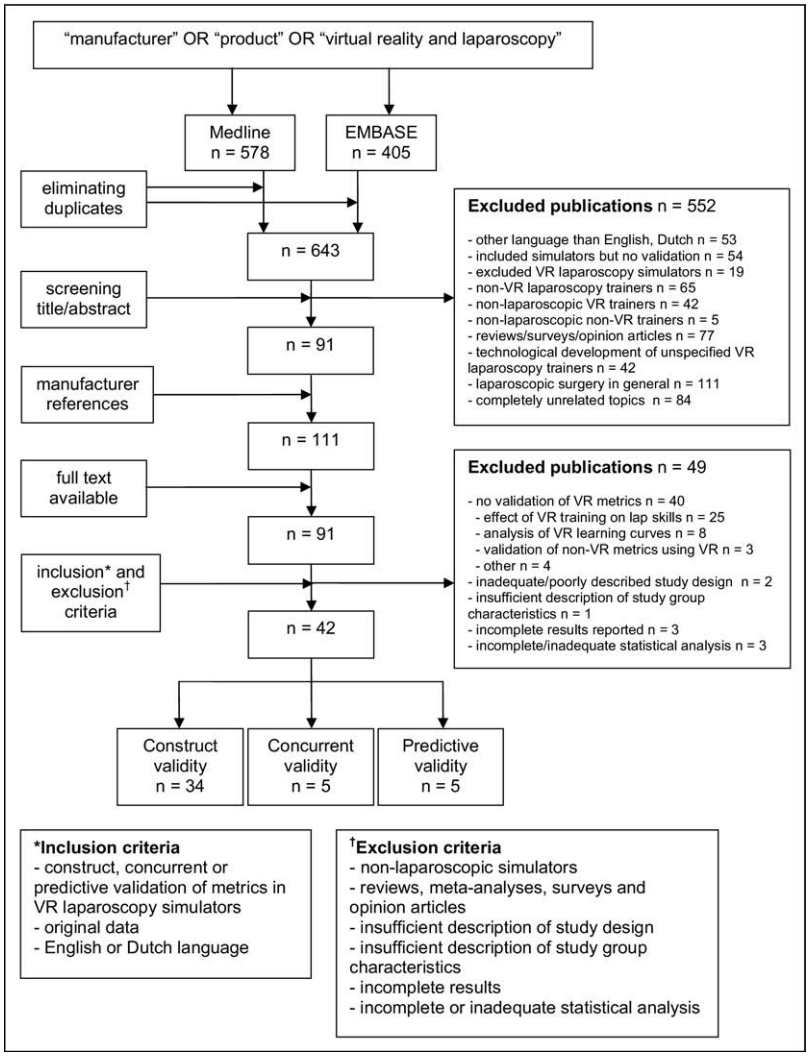

Figure 1 Flow chart depicting identification of relevant publications (search last performed on January 1st, 2008).

ing validation of composite scores, the included metrics or their relative weights differed from the factory default settings.[26,29,34,36,39,42]

The time metric was rated fully construct valid (ie, significant difference in outcome between groups of different training levels in all simulator tasks studied) in all validation studies in which it was assessed for the Haptica ProMIS,[10,11,38] Simbionix Lap Mentor,[22] and Xitact LS 500.[4,37] The time metric was found to display partial validity in publications[8,15,16,18–20,26,29–31,33,35] and no significant validity in 2 publications.[38,39]

The number of errors metric did not reach full construct validity in all the publications pertaining to any of the simulators under study. Of 15 papers reporting on the number of errors metric across 3 simulators, 3 found this parameter to significantly differentiate between groups of participants with different levels of expertise.[9,17,20] Ten papers reported partial validity,[8,19,26,28–32,35,39] while 2 reported no validity.[16,18]

Results for the path length metric were similar to the number of errors metric: full validity was not reached across all studies concerning any type of simulator. Full construct validity was found in 4 publications,[11,24,28,32] partial validity in 13 publications,[6,8,9,11,26,27,29–31,33,35,37,38] and no validity in 1.[38]

Composite scores reached full construct validity in 2 publications,[34,47] partial validity in 9 publications,[4,15,21,23,26,29,34,36,39] and no validity in 4.[7,12,38,42] For none of the simulators the score metric was found to be fully valid across all studies except for Mentice Procedicus KSA, into which a single validation study was performed.[47]

None of the other metrics or scoring systems based reached full construct validity in all publications concerning a specific simulator except for "unnecessary movements" in the Mentice MIST-VR.[17]

Concurrent validity

An overview of the data extracted from the included studies into construct validity is shown in Table 3. Ritter et al[6] tested concurrent validity of the Haptica ProMIS augmented reality simulator in relation to the Fundamentals of

Table 2 Construct validity

Simulator	Publication (first author)	Year	Participants			Study design		Results	
			n	Type	n proc	VR tasks	VR metrics	cv	Significant differences for . . .
Simendo	Verdaasdonk[8]	2007	15	exp	>50	Drop the balls (1 instrument), drop the balls (2 instruments), ring and needle, stretch (easy, difficult), 30° endoscope handling; 3 repetitions for each task	Time	+/−	exp > other groups
			18	int	1-30		Path length	+/−	exp/int > nov/nav for Endoscope and right instrument
			14	nav	1-40		Errors (collisions)	+/−	exp > nov in 2/3 repetitions
			14	nov	0				
	Verdaasdonk[9]	2006	5	exp	>100	Drop the balls (1 instrument), 3 repetitions	Time	+	exp > nov
			20	nov	0		Path length	+	exp > nov
							Errors (collisions)	+	exp > nov
ProMIS	Botden[38]	2007	30	exp	>100	Translocation, suturing; single trial	Time	+	exp > nov, both tasks
			30	nov	0		Path length	+/−	exp > nov, both hands, both tasks (except left hand suturing)
							Smoothness	+	exp > nov, both hands, both tasks
	Ritter[6]	2007	8	exp	<100	FLS peg transfer task; 5 repetitions	Path length	+/−	exp/int > nov (exp vs int: NS)
			8	int	10-100		Smoothness	+/−	exp/int > nov (exp vs int: NS except trial 4)
			44	nov	<10				
	Broe[10]	2006	20	PGY 1-5	?	Tracking/orientation task; 3 repetitions (levels 1-3)	Time	+	Level 2 r^2 = .61, Level 3 r^2 = .81
							n movements	+/−	Level 2 r^2 = .98, level 3 r^2 = .36
	Van Sickle[11]	2005	5	exp	?	Suturing task; 3 repetitions	Time	+	exp > nov
			5	nov	0		Path length	+	exp > nov
							Smoothness	+	exp > nov
MIST-VR	Van Sickle[12]	2007	42	surg	>100	Acquire place, transfer place, traversal, withdraw insert, diathermy task, manipulate diathermy; 5 repetitions for each task	Score	−	Age, prior MIST-VR experience (no difference for years of lap experience, n of lap procedures, no. of lap cholecystectomies)
	Maithel[39]	2006	91	res	?	Manipulate diathermy; 2 repetitions	Time	−	n of advanced cases (training Level, n of basic cases: NS)
				fel	?		Errors	+/−	n of advanced cases (training Level, n of basic cases: NS)
				surg	?		Econ movements	+/−	n of advanced cases (training Level, n of basic cases: NS)
							Econ diathermy	−	
							Score	+/−	
	Hart[7]	2006	15	nov	?	Manipulative daithermy, stretch diathermy, 2-14 repetitions (nov > junior > senior)	Score	−	Score of first repetition (significant results were obtained for best scores but numbers of repetitions varied greatly between groups)
			5	junior	?				
			8	senior	?				
	Avgerinos[15]	2005	12	junior	?	13 different tasks; 3 repetitions	Time	+/−	8 of 13 tasks
			11	int	?		Score	+/−	6 of 13 tasks
			9	senior	?				
	Gallagher[16]	2004	100	exp	>50	Acquire place, transfer place, traversal, withdraw insert, diathermy task, manipulate diathermy; 3 trials, 5 repetitions for each task per trial	Time	+/−	exp > nov
			12	int	<10		Errors	−	
			12	nov	0		Econ movement	+/−	exp > nov only in trial 1
			12	controls	0		Econ diathermy	+/−	nov > int, nov > control
	Grantcharov[17]	2003	8	exp	>100	Acquire place, transfer place, traversal, withdraw insert, diathermy task, manipulate diathermy; 10 trials	Time	+	exp > int > nov
			8	int	15-80		Errors	+	exp > int > nov
			25	nov	<10		Unnec movements	+	exp > int > nov
	Gallagher[18]	2002	12	exp	>100	Acquire place, transfer place, traversal, withdraw insert, diathermy task, manipulate diathermy; 10 trials	Time	+/−	? > ? > ?
			12	int	<10		Errors	−	
			12	nov	0		Econ movements	+	exp > int > nov
							Econ diathermy	−	
	Gallagher[19]	2001	12	exp	>50	Acquire place, transfer place, traversal, withdraw insert, diathermy task, manipulate diathermy; 5 repetitions	Time	+/−	exp > int, exp > nov (int > nov: NS)
			12	int	<10		Errors	+/−	exp > int
			12	nov	0		Econ movements	+/−	exp > int, exp > nov (int > nov: NS)
							Econ diathermy	−	

(continued on next page)

Table 2 *(continued)*

Simulator	Publication (first author)	Year	Participants n	Participants Type	Participants n proc	Study design VR tasks	VR metrics	Results cv	Results Significant differences for . . .
	McNatt[20]	2001	3	exp	>250	Object acquisition, target traversal, target	Time	+/−	exp > nov in 2/3 tasks
			5	nov	0-10	manipulation and diathermy; 7-10 trials	Errors	+	exp > nov in all tasks
							Econ movements	+/−	exp > nov in 2/3 tasks
Procedicus KSA	Felländer-Tsai[47]	2004	10	exp	?	Instrument navigation; 3 repetitions	Score	+	exp > nov
			10	nov	0				
Lap Mentor	Zhang[21]	2007	9	res	10-20	9 basic skills tasks, pre- and post-test	Score	+/−	res > mst in 3/18 comparisons
			9	mst	0				res > nov in 13/18 comparisons
			9	nov	0				ms > nov in 13/18 comparisons
	Yamaguchi[22]	2007	16	exp	>50	Eye-hand coordination task; single trial	Time	+	exp > nov
			15	nov	<10		Econ movements	+/−	exp > nov left hand (right hand: NS)
							n movements	+/−	exp > nov left hand (right hand: NS)
							Speed	+/−	exp > nov left hand (right hand: NS)
	McDougall[23]	2006	30	exp 1	>30/y	9 basic skills tasks; single trial	Score	+/−	exp 1/res > exp 2, exp 2 > ms
			26	exp 2	<30/y				
			24	res	?				
			23	ms	?				
LapSim	Botden[38]	2007	30	exp	>100	Translocation, suturing; single trial	Time	−	
			30	nov	0		Path length	−	
							Damage	+/−	exp > nov in suturing task
							Score	−	
	Ahlberg[24]	2007	5	exp	>300	Grasping, lift grasp, cutting right, cutting left,	Time	+	exp > res
			13	res	?	and clip application; single trial	Path length	+	exp > res
							Angular path	+	exp > res
							Damage	−	
							Max damage	−	
	Van Dongen[26]	2007	16	exp	>100	Camera navigation, instrument navigation,	Time	+/−	exp/res > nov (exp vs res: NS)
			16	res	10-100	coordination, grasping, lifting and grasping,	Path length	+/−	exp/res > nov (exp vs res: NS)
			16	nov	0	cutting, and clipping and cutting; 3 levels	Errors	+/−	exp/res > nov (exp vs res: NS)
							Overall score*	+/−	exp/res > nov (exp vs res: NS)
	Aggarwal[27]	2006	8	exp	>100	Instrument navigation; single session	Time	+	exp > int > nov in all 3 sessions
			8	int	20-50	salpingectomy Procedural module; 3	Path length	+/−	exp > int > nov in navigation, 2nd salpingectomy session
			7	nov	<10	repetitons, 2 sessions	Blood loss	+/−	exp > int > nov in 2nd salpingectomy session
	Aggarwal[28]	2006	10	exp	>100	7 basic tasks; 10 sessions (first two used for	Time	+	exp > nov
			10	nov	<10	construct validity assessment)	Path length	+	exp > nov
							Errors	+/−	exp > nov for 1/7 tasks
			9	exp	>100	Dissection of Calot Triangle procedural module;	Time	+	exp > nov
			11	nov	<10	exp 2 repetitions, nov 10 repetitions (first	Path length	+	exp > nov
						two used for construct validity assessment)	Angular path	+	exp > nov
							Blood loss	+	exp > nov
							Dissected volume	+	exp > nov
	Larsen[29]	2006	10	exp	>100/y	Lfting and grasping, cutting, clipping, ectopic	Time	+/−	exp > int/nov (int vs nov: NS)
			10	int	20-60/2y	pregnancy Procedural module; 10 sessions	Path length	+/−	exp > int/nov (int vs nov: NS)
			10	nov	0	(first two used for construct validity	Angular path	+/−	exp > int/nov (int vs nov: NS, NS in ectopic pregnancy module)
						assessment)	Errors	+/−	exp > int/nov (int vs nov: NS in 1/5 types of errors)
							Score*	+/−	exp > int/nov (int vs nov: NS)
	Hassan[30]	2006	19	int	30-50	Clip application; 2 sessions, before and after	Time	+/−	int > nov in first session
			48	nov	<10	training	Path length	+/−	int > nov in first session
							Angular path	+/−	int > nov in first session
							Errors	+/−	int > nov in both sessions, except for blood loss

(continued on next page)

Table 2 *(continued)*

Simulator	Publication (first author)	Year	Participants n	Participants Type	Participants n proc	Study design VR tasks	Study design VR metrics	Results cv	Results Significant differences for . . .
	Woodrum[31]	2006	5	exp	?	Coordination, instrument navigation, grasping,	Time	+/−	exp > res > nov in 4/6 tasks, exp > res/nov in 1/6 tasks
			20	res	?	lifting and grasping, cutting, clip applying;	Path length	+/−	exp > res > nov or exp > res/nov in 2/5 tasks
			9	nov	0	10 repetitions	Angular path	+/−	exp > res > nov in 1/6 tasks, exp > res/nov in 1/6 tasks
							Errors	+/−	exp > res > nov in 2/6 tasks, exp/res > nov in 2/6 tasks
	Eriksen[32]	2005	10	exp	>100	Camera navigation, instrument navigation,	Time	+	exp > nov
			14	nov	<10	coordination, grasping, lifting and grasping,	Path length	+	exp > nov
						cutting, clip applying; 3 repetitions	Angular path	+	exp > nov
							Errors	+/−	exp > nov in 7/19 comparisons
	Langelotz[33]	2005	54	exp	>50	Navigation, coordination, grasping, cutting,	Time	+/−	exp > int in 4/5 tasks and overall
			61	int	<50	clipping; single trial	Path length	+/−	exp > int in 1/5 tasks (other tasks ?)
							Angular path	+/−	exp > int in 1/5 tasks (other tasks ?)
							Damage	+/−	exp > int in 1/5 tasks (other tasks ?)
	Sherman[34]	2005	7	exp	≥50	Grasping, cutting, clipping; 7-12 repetitions	Time-error score*	+	exp > res > nov
			10	res	<25	(first and last repetitions used for construct	Motion score*	+/−	exp > res > nov in first repetition, last repetition: NS
			7	nov	0	validity assessment)			
	Ro[42]	2005	13	int	>30	Navigation, coordination, grasping, lifting and	Performance score*	−	nov > int in 6/8 tasks
			16	nov	0	grasping, cutting, clipping, suturing,			
						dissection	Efficiency score*	−	nov > int in 6/8 tasks
	Duffy[35]	2005	7	exp	?	Camera navigation, instrument navigation,	Time	+/−	exp > nov in 5/5 tasks (res ?)
			37	res	?	coordination, grasping, lifting and grasping,	Left path length	+/−	exp > nov in 2/4 tasks (other tasks, res ?)
			10	nov	?	cutting, clip applying, suturing	Errors	+/−	exp > nov in 3/5 tasks (other tasks, res ?)
EndoTower	Stefanidis[36]	2007	90	nov-exp	(0->251)	1 task; 2 repetitions	Score*	+/−	n of lap cholecystecomies, n of lap cases, frequency of angled scope use (training level: NS, lap fellowship: NS)
	Maithel[39]	2006	91	nov-exp	?	1 task; 2 repetitions	Score*	+/−	n of advanced lap cases, lap fellowship yes/no (training level, n of basic cases: NS)
LS500	Rosenthal[37]	2007	150	bc	?	Clipping and cutting of the cystic artery and	Time	+	ic > bc
			157	ic	?	cystic duct; 1-3 repetitions	Path length	+/−	ic > bc right hand (left hand: NS)
	Schijven[4]	2003	37	exp	>100	Clipping and cutting of the cystic artery and	Time	+	exp > nov
			37	nov	0	cystic duct; 3 repetitions	Score	+/−	exp > nov in 2nd and 3rd repetition

n proc = total number of procedures performed; cv = construct validity; nov = novice; exp = expert; res = resident; fel = advanced laparoscopy fellow; surg = practicing laparoscopic surgeon; int = intermediate; bc = basic course participant; ic = intermediate course participant; nav = endoscope navigation experience; mst = medical student; /y, /2y = per year, per 2 years; Econ = economy; Unnec = unnecessary; + = full construct validity (for all comparisons between groups in all tasks); +/− = partial construct validity (for some comparisons between groups/in some tasks); − = no construct validity; > = ". . . achieved better results than . . ."; NS = not significant; ? = not reported/unknown.

*Parameters or relative weights in sum score differ from factory default settings.

Table 3 Concurrent validity

Publication (first author)	Year	Comparison	Participants		Study design				Results
			n	Type	VR tasks	VR metrics	Reference tasks	Reference metrics	Pearson correlations
Ritter[6]	2007	ProMIS vs FLS metrics	8 8 44	exp int nov	Peg transfer, 5 repetitions	Smoothness Path length	Same as ProMIS	Score (based on time, errors)	Path length:.78 (nov), .5 (int: NS), .86 (exp). Smoothness: .94 (nov), .98 (int), .99 (exp).
Madan[46]	2003	MIST-VR vs box	16	nov	Peg placement L Peg placement R	Time Econ mov Errors	Peg placement L Peg placement R Peg transfer L to R	Time Errors	Single handed: no correlation. Two handed: correlation between time (box) and time/econ mov (VR), no correlation for errors.
Newmark[48]	2007	LapSim vs box	34	nov	Coordination lifting and grasping handling intestines	Time Damage	Peg transfer Dropping beans Passing rope	Time Errors	Out of 36 combinations of VR and box metrics 9 correlated.
Schijven[49]	2004	LS500 vs aptitude test battery	33	nov	Clip-and-cut task	Score (based on time, errors)	Abstract Reasoning, Space Relations, Gibson Spiral Maze, Crawford Small Parts Dexterity Tester	Score	Abstract reasoning test was the only test correlating significantly to Xitact test outcome.
Haluck[50]	2002	EndoTower vs aptitude test battery	25	Nov	Identifying randomly placed arrows, 3 repetitions	Score (based on time, errors)	PicSOr, Card Rotation Cube Comparison, Map Plan tests	Score	PicSOr: $r = .591$, Card Rotation: $r = .296$, Cube Comparison: $r = .354$, Map Plan: $r = .392$, PicSOr + Card Rotation: $r = .658$, PicSOr + Cube comparison: $r = .701$, PicSOr + Map Plan: $r = .708$

nov = novice; exp = expert; int = intermediate; NS = not significant; L = left; R = right.

Laparoscopic Surgery (FLS) program score. They included 60 volunteers in a trial consisting of up to 5 repetitions of the FLS Peg Transfer task in the ProMIS simulator. Subjects were stratified into 3 groups: experienced (>100 laparoscopic procedures, n = 8), novices (<10 laparoscopic procedures, n = 44), and intermediates (n = 8). The simulator recorded execution time, instrument path length, and instrument smoothness metrics. The performances were videotaped and the FLS score was calculated based on execution time and number of errors. Using Pearson's test for linear correlation, path length and smoothness metrics showed strong relationships with the FLS scores.

Madan et al[46] compared box trainer and MIST-VR performance scores of 16 students who had no previous laparoscopic experience. For the box performances, time and number of errors were recorded for tasks performed with the dominant, nondominant, and both hands. The MIST-VR task was performed with both the dominant and nondominant hand. The simulator recorded time, economy of movement, and errors. For the 1-handed tasks, these metrics did not correlate with time and errors on the box trainer. For the 2-handed task on the box trainer, time correlated with time and economy of movement on the VR trainer, but there was no correlation between errors on the box and VR trainers.

Newmark et al[48] compared time to completion and tissue damage scores on the LapSim with completion time and errors on a box trainer for 34 third-year medical students. Three different tasks were analyzed on the LapSim and 3 on the box trainer. Pearson correlations were calculated for 36 combinations of box and VR metrics; only 9 showed strong correlations.

Schijven et al[49] compared Xitact LS 500 simulator performance and results from 4 aptitude tests (Crawford Small Parts Dexterity Tester, Abstract Reasoning test, Space Relations test, and Gibson Spiral Maze) for 33 laparoscopic novices. The simulator test consisted of a "clipping-and-cutting" task for which the simulator calculated a composite score based on task time and several types of errors. The Abstract Reasoning test was the only test correlating significantly to the Xitact test results.

Haluck et al[51] also compared VR metrics (Endotower) with mental aptitude tests (PicSOr, Card Rotation, Cube Comparison, and Map Plan tests). All participants were laparoscopic novices (n = 25). The Endotower score, based on time and number of errors, correlated well with the PicSOr test but weakly with the other tests. Combinations between PicSOr and the other tests showed increased correlations in a multiple regression model.

Predictive validity

The literature search yielded 16 "VR to OR" publications, in which the effect of VR training on in vivo laparoscopic performance is assessed.[7,25,51–64] Five publications reported predictive validity. Study characteristics and results are presented in Table 4.

Ahlberg et al[51] studied the learning curves for laparoscopic fundoplication in 12 pairs of laparoscopic "masters" and "pupils"; the latter were surgeons with laparoscopic experience in general but not in fundoplication specifically. Each pupil performed 20 consecutive fundoplications, which were all videotaped. For each master/pupil pair, videos of 1 operation by the master and 5 by the pupil were rated by independent reviewers. At the start of the trial the pupils were tested in the Procedicus MIST simulator. The following metrics were recorded: economy in movements, economy in diathermy used, number of errors, and time used for each task in the simulator. There was only a weak correlation between the MIST-VR results and the scores from the first and last procedures, $r = .25$ and $r = .15$, respectively.

Another study by Ahlberg et al[55] compared laparoscopic performance of VR-trained (n = 14) and nontrained subjects (n = 15) in a porcine model. The participants were fourth-year medical students with no laparoscopic experience. The surgical task in the porcine model was a simulated appendectomy. The performances were videotaped and scored by 2 independent observers on a 0–2 scale for 5 parts of the procedure. Comparison of the video scores with scores from the most complex MIST-VR task showed a weak correlation ($r = .33$). Exclusion of 2 statistical outliers yielded a stronger correlation ($r = .64$).

Madan et al[52] recruited 32 medical students with no previous laparoscopic experience to perform 2 operative tasks in a porcine model. The performances were timed and scored on a subjective 0–100 scale. Subsequently, the subjects repeated 1 task on the MIST-VR trainer several times. The simulator recorded time and economy of movement. Of the 16 possible relationships between performance metrics in the OR and on the simulator, 11 showed a statistically significant correlation.

Grantcharov et al[57] included 14 surgical residents with limited laparoscopic experience (<10 cholecystectomies). The participants performed all 6 MIST-VR tasks on the first and third days of a laparoscopic skills course. Economy of movement for both hands and number of errors were recorded. On the second day the subjects performed a laparoscopic cholecystectomy in a pig. The procedures were scored by observers for economy of movement and errors. Using Spearman's test correlations were established: for errors, the porcine score and 3 of 6 VR tasks correlated. For economy of movement, correlations for the in vivo score and 5 of 6 VR tasks (right hand) and 1 VR task (left hand) were observed.

Hart et al[7] compared MIST-VR performance with gynecologic procedures performed in an ovine model: salpingotomy, salpingectomy, and tubal clipping. The participants were medical students (n = 15), junior doctors (n = 6), and senior doctors (n = 8). The participants performed varying numbers of MIST-VR training sessions during a 2-month period: they performed the 3 ovine procedures once before and once and after the simulator training. The VR metric

Table 4 Predictive validity

Publication (first author)		Simulator	Participants		Study design				Results
			N	Type	VR tasks	VR metrics	OR tasks	OR metrics	
Ahlberg[51]	2005	Procedicus MIST	12	lap surgeons training fundoplication	Task 6 (most coplex task)	Time Econ mov Econ diath Errors	20 consecutive laparoscopic fundoplications	0-3 score for 7 parts of operation	*Pearson correlations* First operation .25, last operation .15.
Madan[52]	2005	MIST-VR	32	nov	Acquire and place task	Time (L/R) Econ mov (L/R)	Porcine bowel measuring, bowel placement in bag	Time, subjective 0-100 score	*Pearson correlations* Statistically significant correlation for 11 of 16 possible relations between VR and OR metrics.
Ahlberg[54]	2002	MIST-VR	14	nov	Task 6 (most coplex task)	Score	Porcine simulated lap appendectomy	0-2 score for 5 parts of operation	*Regression analysis* With outliers: .33 Without 2 outliers: .64
Grantcharov[57]	2001	MIST-VR	14	res	6 tasks	Econ mov (L/R) Errors	Porcine lap cholecystectomy	Econ movement Errors	*Spearman's test* Errors: correlation for porcine score and 3/6 VR tasks Econ mov: correlationfor porcine score and 5/6 VR tasks (R) and 1 VR task (L)
Hart[7]	2006	MIST-VR	22-29	nov, junior surgeons, senior surgeons	Manipulative daithermy, stretch diathermy	Score	2 ovine salpingotomies, salpingectomies, tubal clippings	Percentile rank (based on time, Score)	*Bivariate correlation and regression modeling* Statistically significant correlation for 2 of 12 possible relations between VR and OR metrics

nov = novice; int = intermediate; res = resident; Econ = economy; L = left hand; R = right hand.

used in this study was "score," while the OR performance was rated in percentile ranks based on a score composed of time taken to complete the task and various penalty scores. Of 12 possible combinations between the baseline MIST-VR scores and operative percentile ranks, only 2 showed a statistically significant correlation: the baseline MIST-VR manipulative diathermy score correlated with the pretraining salpingotomy percentile rank and the baseline MIST-VR stretch diathermy correlated with pretraining tubal clipping. Correlations between MIST-VR best scores and OR performance were not included in this review due to the great variance in numbers of training sessions between participants.

Comments

Systematic reviews should feature critical appraisal of the quality of the included studies.[65] Carter et al[66] conducted a review of validation evidence for surgical simulators similar to this review. Levels of evidence were awarded to the included studies using a grading system based on "the principle of evidence-based guideline development," similar to the Oxford Centre for Evidence-Based Medicine Levels of Evidence.[67] Full-text publications were considered level 2b/c evidence, ie, "nonrandomized trials, comparative research," while abstracts lacking the detail required for judgment of their quality were rated level 4 ("expert opinions"). As this grading system was designed to rate clinical evidence, it is not ideally suited for grading validation studies. Evidence level 1b ("randomized controlled trial of good quality") can by definition never be achieved by a construct validation study because random allocation of different interventions is irrelevant to its study design, which is more similar to a diagnostic study. The same applies to studies into concurrent and predictive validity. If this grading system were to be applied to the publications included in this review, all articles would be awarded level 2. Unfortunately, guidelines more suitable for the critical appraisal of validation studies could not be identified by the authors. This review therefore lacks a formal assessment of the methodological quality of the included publications.

The included publications concerning construct validation showed considerable heterogeneity with respect to study design characteristics such as the number of participants, the level of participants' experience, the way levels of experience were defined, the metrics under study, and the types of tasks used for their validation. This heterogeneity may account for some of the differences in results between studies into the same simulators and metrics.

Five publications into concurrent validity of simulator metrics were identified.[6,46,48–50] In these publications, simulator metrics were compared with other, "established" assessment methods for surgical skills. The Haptica ProMIS metrics correlated well with the Fundamentals of Laparoscopic Surgery normalized score.[6] Comparisons between VR metrics and box trainer metrics[46,48] and mental aptitude tests[49,50] were less convincing.

The literature search yielded 16 "VR to OR" publications, in which the effect of VR training on in vivo laparoscopic performance was assessed.[7,25,51–64] The "VR to OR" study design allows a comparison to be made between the VR performance metrics and assessment of real life (in vivo) surgical performance, thus establishing the predictive validity of the simulator. In vivo surgical proficiency is, of course, the main goal of any surgical training program. One might argue that, compared with predictive validity construct validity, the ability to differentiate between individuals with (supposedly) different skill levels in an in vitro setting, is of lesser importance. Similarly, predictive validity can be regarded as being superior to concurrent validity; how well a VR simulator's metrics correspond to those of a validated box trainer is a "secondary outcome measure" compared with being able to predict real life surgical performance.

It is therefore disappointing that of 16 "VR to OR" publications, only 5 presented data on predictive validity.[7,51,52,55,57] None of these studies used a validated system such as GOALS[68] for the skills assessment during the in vivo tasks. The only study in which operations in human patients were employed for assessment of laparoscopic (as opposed to ovine or porcine models) skill showed the weakest correlation between VR and OR performance.[52] All studies were performed using either the Mentice MIST-VR of Procedicus MIST simulator; similar studies into other commonly used systems were not identified.

A recent systematic review[69] has demonstrated VR training has the potential to supplement the standard "apprenticeship" form of training. In many teaching institutions around the world, VR laparoscopy simulators are now available and often integrated into the surgical curriculum. With careful selection of the simulator system, tasks and metrics, it seems feasible to use VR systems for the assessment of laparoscopic skills. The results tables in this review may serve as a guide to choosing the simulator, tasks, and metrics for this purpose. However, these assessments may not necessarily predict in vivo surgical performance. More research into predictive validity of simulator metrics is needed, preferably using a standardized methodology and validated assessment systems for the in vivo performances.

Conclusion

Assessment of laparoscopic skills utilizing VR laparoscopy metrics seems feasible if the VR system, tasks and metrics are carefully selected. However, such assessments will only provide insight in the relative skill levels of trainees and will not yet predict real life surgical proficiency. More research is needed to establish predictive validity of VR laparoscopy simulator metrics.

References

1. Scott DJ, Bergen PC, Rege RV, et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? J Am Coll Surg 2000;191:272–83.
2. Nelson AA. Research design: measurement, reliability, and validity. Am J Hosp Pharm 1980;37:851–7.
3. Wanzel KR, Ward M, Reznick RK. Teaching the surgical craft: from selection to certification. Curr Probl Surg 2002;39:573–659.
4. Schijven M, Jakimowicz J. Construct validity: experts and novices performing on the Xitact LS 500 laparoscopy simulator. Surg Endosc 2003;17:803–10.
5. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 2003;17:1525–9.
6. Ritter EM, Kindelan TW, Michael C, et al. Concurrent validity of augmented reality metrics applied to the Fundamentals of Laparoscopic Surgery (FLS). Surg Endosc 2007;21:1441–5.
7. Hart R, Doherty DA, Karthigasu K, et al. The value of virtual reality-simulator training in the development of laparoscopic surgical skills. J Minim Invasive Gynecol 2006;13:126–33.
8. Verdaasdonk EG, Stassen LP, Schijven MP, et al. Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. Surg Endosc 2007;21:1406–12.
9. Verdaasdonk EG, Stassen LP, Monteny LJ, et al. Validation of a new basic virtual reality simulator for training of basic endoscopic skills: the SIMENDO. Surg Endosc 2006;20:511–8.
10. Broe D, Ridgway PF, Johnson S, et al. Construct validation of a novel hybrid surgical simulator. Surg Endosc 2006;20:900–4.
11. Van Sickle KR, McClusky DA 3rd, Gallagher AG, et al. Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surg Endosc 2005;19:1227–31.
12. Van Sickle KR, Ritter EM, McClusky DA 3rd, et al. Attempted establishment of proficiency levels for laparoscopic performance on a national scale using simulation: the results from the 2004 SAGES Minimally Invasive Surgical Trainer-Virtual Reality (MIST-VR) Learning Center study. Surg Endosc 2007;21:5–10.
13. Datta V, Bann S, Aggarwal R, et al. Technical skills examination for general surgical trainees. Br J Surg 2006;93:1139–46.
14. Brunner WC, Korndorffer JR Jr, Sierra R, et al. Determining standards for laparoscopic proficiency using virtual reality. Am Surg 2005;71:29–35.
15. Avgerinos DV, Goodell KH, Waxberg S, et al. Comparison of the sensitivity of physical and virtual laparoscopic surgical training simulators to the user's level of experience. Surg Endosc 2005;19:1211–5.
16. Gallagher AG, Lederman AB, McGlade K, et al. Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality (MIST-VR) using criteria levels based on expert performance. Surg Endosc 2004;18:660–5.
17. Grantcharov TP, Bardram L, Funch-Jensen P, et al. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. Am J Surg 2003;185:146–9.
18. Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surg Endosc 2002;16:1746–52.
19. Gallagher AG, Richie K, McClure N, et al. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 2001;25:1478–83.
20. McNatt SS, Smith CD. A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. Surg Endosc 2001;15:1085–9.
21. Zhang A, Hunerbein M, Dai Y, et al. Construct validity testing of a laparoscopic surgery simulator (Lap Mentor): evaluation of surgical skill with a virtual laparoscopic training simulator. Surg Endosc 2008;22:1440–4.
22. Yamaguchi S, Konishi K, Yasunaga, et al. Construct validity for eye-hand coordination skill on a virtual reality laparoscopic surgical simulator. Surg Endosc 2007;21:2253–7.
23. McDougall EM, Corica FA, Boker JR, et al. Construct validity testing of a laparoscopic surgical simulator. J Am Coll Surg 2006;202:779–87.
24. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg 2007;193:797–804.
25. Petrin P, Baggio E, Spisni R, et al. Use of virtual reality simulator in the training of postgraduated surgical residents. Ann Ital Chir 2006;77:465–8.
26. van Dongen KW, Tournoij E, Van Der Zee DC, et al. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? Surg Endosc 2007;21:1413–7.
27. Aggarwal R, Tully A, Grantcharov T, et al. Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy. Br J Obstet Gynaecol 2006;113:1382–7.
28. Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg 2006;244:310–4.
29. Larsen CR, Grantcharov T, Aggarwal R, et al. Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. Surg Endosc 2006;20:1460–6.
30. Hassan I, Maschuw K, Rothmund M, et al. Novices in surgery are the target group of a virtual reality training laboratory. Eur Surg Res 2006;38:109–13.
31. Woodrum DT, Andreatta PB, Yellamanchilli RK, et al. Construct validity of the LapSim laparoscopic surgical simulator. Am J Surg 2006;191:28–32.
32. Eriksen JR, Grantcharov T. Objective assessment of laparoscopic skills using a virtual reality stimulator. Surg Endosc 2005;19:1216–9.
33. Langelotz C, Kilian M, Paul C, et al. Virtual reality laparoscopic simulator reflects clinical experience in German surgeons. Langenbecks Arch Surg 2005;390:534–7.
34. Sherman V, Feldman LS, Stanbridge D, et al. Assessing the learning curve for the acquisition of laparoscopic skills on a virtual reality simulator. Surg Endosc 2005;19:678–82.
35. Duffy AJ, Hogle NJ, McCarthy H, et al. Construct validity for the LapSim laparoscopic surgical simulator. Surg Endosc 2005;19:401–5.
36. Stefanidis D, Haluck R, Pham T, et al. Construct and face validity and task workload for laparoscopic camera navigation: virtual reality versus videotrainer systems at the SAGES Learning Center. Surg Endosc 2007;21:1158–64.
37. Rosenthal R, Gantert WA, Hamel C, et al. Assessment of construct validity of a virtual reality laparoscopy simulator. J Laparoendosc Adv Surg Tech A 2007;17:407–13.
38. Botden SM, Buzink SN, Schijven MP, et al. 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 2007;31:764–72.
39. Maithel S, Sierra R, Korndorffer J, et al. Construct and face validity of MIST-VR, Endotower, and Celts: are we ready for skills assessment using simulators? Surg Endosc 2006;20:104–12.
40. Felländer-Tsai L, Kjellin A, Wredmark T, et al. Basic accreditation for invasive image-guided intervention: a shift of paradigm for high technology education, embedding performance criterion levels in advanced medical simulators in a modern educational curriculum. The Journal on Information Technology in Healthcare 2004;3:165–73.
41. Hassan I, Gerdes B, Koller M, et al. Clinical background is required for optimum performance with a VR laparoscopy simulator. Comput Aid Surg 2006;11:103–6.
42. Ro CY, Toumpoulis IK, Ashton RC Jr, et al. The LapSim: a learning environment for both experts and novices. Stud Health Technol Inform 2005;111:414–7.
43. Chaudhry A, Sutton C, Wood J, et al. Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human-computer interface. Ann R Coll Surg Engl 1999;81:281–6.

44. Heinrichs WL, Lukoff B, Youngblood P, et al. Criterion-based training with surgical simulators: proficiency of experienced surgeons. JSLS 2007;11:273–302.
45. Strom P, Kjellin A, Hedman L, et al. Validation and learning in the Procedicus KSA virtual reality surgical simulator. Surg Endosc 2003; 17:227–31.
46. Madan AK, Frantzides CT, Shervin N, et al. Assessment of individual hand performance in box trainers compared to virtual reality trainers. Am Surg 2003;69:1112–4.
47. Fellander-Tsai L, Stahre C, Anderberg B, et al. Simulator training in medicine and health care. A new pedagogic model for good patient safety. Lakartidningen 2001;98:3772–6.
48. Newmark J, Dandolu V, Milner R, et al. Correlating virtual reality and box trainer tasks in the assessment of laparoscopic surgical skills. Am J Obstet Gynecol 2007;197:e1–4.
49. Schijven MP, Jakimowicz JJ, Carter FJ. How to select aspirant laparoscopic surgical trainees: establishing concurrent validity comparing Xitact LS 500 index performance scores with standardized psychomotor aptitude test battery scores. J Surg Res 2004;121:112–9.
50. Haluck RS, Gallagher AG, Satava RM, et al. Reliability and validity of Endotower, a virtual reality trainer for angled endoscope navigation. Stud Health Technol Inform 2002;85:179–84.
51. Ahlberg G, Kruuna O, Leijonmarck CE, et al. Is the learning curve for laparoscopic fundoplication determined by the teacher or the pupil? Am J Surg 2005;189:184–9.
52. Madan AK, Frantzides CT, Sasso LM. Laparoscopic baseline ability assessment by virtual reality; Laparoendosc J Surg Tech A 2005;15: 13–7.
53. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004;91:146–50.
54. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002;236:458–63.
55. Ahlberg G, Heikkinen T, Iselius L, et al. Does training in a virtual reality simulator improve surgical performance? Surg Endosc 2002; 16:126–9.
56. Hamilton EC, Scott DJ, Fleming JB, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. Surg Endosc 2002;16:406–11.
57. Grantcharov TP, Rosenberg J, Pahle E, et al. Virtual reality computer simulation. Surg Endosc 2001;15:242–4.
58. Andreatta PB, Woodrum DT, Birkmeyer JD, et al. Laparoscopic skills are improved with LapMentor training: results of a randomized, double-blinded study. Ann Surg 2006;243:854–60.
59. Aggarwal R, Ward J, Balasundaram I, et al. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. Ann Surg 2007;246:771–9.
60. Youngblood PL, Srivastava S, Curet M, et al. Comparison of training on two laparoscopic simulators and assessment of skills transfer to surgical performance. J Am Coll Surg 2005;200:546–51.
61. Hyltander A, Liljegren E, Rhodin PH, et al. The transfer of basic skills learned in a laparoscopic simulator to the operating room. Surg Endosc 2002;16:1324–8.
62. Ganai S, Donroe JA, St. Louis MR, et al. Virtual-reality training improves angled telescope skills in novice laparoscopists. Am J Surg 2007;193:260–5.
63. Schijven MP, Jakimowicz JJ, Broeders IA, et al. The Eindhoven laparoscopic cholecystectomy training course—improving operating room performance using virtual reality training: results from the first E.A.E.S. accredited virtual reality trainings curriculum. Surg Endosc 2005;19:1220–6.
64. Kimura T, Kawabe A, Suzuki K, et al. Usefulness of a virtual reality simulator or training box for endoscopic surgery training. Surg Endosc 2006;20:656–9.
65. Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. Ann Intern Med 1997;127:531–7.
66. Carter FJ, Schijven MP, Aggarwal R, et al. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 2005;19: 1523–32.
67. Centre for Evidence-Based Medicine Levels of Evidence. 2001. Available at: http://www.cebm.net/index.aspx?o=1025. Accessed April 4, 2008.
68. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005;190:107–13.
69. Gurusamy KS, Aggarwal R, Palanivelu L, et al. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2009;1:CD006575.