Surg Endosc (2013) 27:1468–1477
DOI 10.1007/s00464-012-2631-7

REVIEWS

Is motion analysis a valid tool for assessing laparoscopic skill?

John D. Mason · James Ansell · Neil Warren · Jared Torkington

Received: 1 June 2012 / Accepted: 21 September 2012 / Published online: 12 December 2012

Abstract

Background The use of simulation for laparoscopic training has led to the development of objective tools for skills assessment. Motion analysis represents one area of focus. This study was designed to assess the evidence for the use of motion analysis as a valid tool for laparoscopic skills assessment.

Methods Embase, MEDLINE and PubMed were searched using the following domains: (1) motion analysis, (2) validation and (3) laparoscopy. Studies investigating motion analysis as a tool for assessment of laparoscopic skill in general surgery were included. Common endpoints in motion analysis metrics were compared between studies according to a modified form of the Oxford Centre for Evidence-Based Medicine levels of evidence and recommendation.

Results Thirteen studies were included from 2,039 initial papers. Twelve (92.3 %) reported the construct validity of motion analysis across a range of laparoscopic tasks. Of these 12, 5 (41.7 %) evaluated the ProMIS Augmented Reality Simulator, 3 (25 %) the Imperial College Surgical Assessment Device (ICSAD), 2 (16.7 %) the Hiroshima University Endoscopic Surgical Assessment Device (HUESAD), 1 (8.33 %) the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT) and 1 (8.33 %) the Robotic and Video Motion Analysis Software (ROVIMAS). Face validity was reported by 1 (7.7 %) study each for ADEPT and ICSAD. Concurrent validity was reported by 1 (7.7 %) study each for ADEPT, ICSAD and ProMIS. There was no evidence for predictive validity.

Conclusions Evidence exists to validate motion analysis for use in laparoscopic skills assessment. Valid parameters are time taken, path length and number of hand movements. Future work should concentrate on the conversion of motion data into competency-based scores for trainee feedback.

Keywords Quality Control · Surgical < technical · Education

Subjective methods of trainee assessment are no longer adequate for surgical training [1]. Reduced working hours [2, 3], increased demands from the political sector [4] and financial pressures [5] mean that more objective measures are required. Surgical simulation is an effective tool for training and assessment. Simulators can reduce learning curves outside the operating theatre in a pressure-free environment, without requiring formal supervision [6]. Studies show that skills acquired during simulation training are transferable to the operating room [7]. Simulation in laparoscopic training refers to a wide range of devices from simple box trainers [8], cadaveric models [9], live animals

J. D. Mason
Cardiff University School of Medicine, Welsh Institute of Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, UK
e-mail: masonjd85@gmail.com

J. Ansell (✉)
Royal College of Surgeons of England Research Fellow, Welsh Institute of Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, UK
e-mail: ansellj@cf.ac.uk

N. Warren
Postgraduate Deanery Wales, Welsh Institute of Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, UK

J. Torkington
Department of Colorectal Surgery, University Hospital of Wales, Heath Park, Cardiff, UK

[10], to complex virtual reality (VR) systems (e.g. MIST-VR®, LapSim® ProMIS™, and LapMentor™) [11–14]. This has led to the development of simulator assessment tools which include motion analysis.

Motion analysis allows assessment of surgical dexterity using parameters that are extracted from movement of the hands or laparoscopic instruments [15]. Several different motion analysis systems have been developed (Table 1). This can be inbuilt within a simulator (e.g. ProMIS™) or as a separate device, enabling flexible use (e.g. Imperial College Surgical Assessment Device, ICSAD) [16]. Objective assessment of laparoscopic skill could be carried out using motion analysis if endpoints for each parameter are quantified according to pre-defined levels of experience. The conversion of motion analysis data into competency-based scores or indexes could provide a valuable source of trainee feedback [17]. This is an automatic and instant process [18]. Feedback could be useful on two levels, firstly by providing a quantitative index to define varying levels of experience, which trainees can work towards. Secondly, it could serve as evidence of professional development that is assessed at annual progress reviews. Before motion analysis can be used to assess laparoscopic competence, the technology and metrics measured must first be validated [19].

Validation of any new method for training or assessment is a critical step [20]. This is the extent to which an instrument measures what it was designed to measure [21, 22]. The process should begin by defining a "construct", which defines the underlying trait for which a new training tool is designed [20]. The more forms of validity (Table 2) that are demonstrated, the stronger the overall argument [20].

The aims of this systematic review are to provide an overview of the different motion analysis technologies available for the assessment of laparoscopic skill, and to identify the evidence for their validity.

Methods

Data resources and search criteria

A systematic review was performed according to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines [23]. The literature search was conducted using the following databases: Embase Classic + Embase (1947 to 2011 week 38), MEDLINE (1947 to present) and PubMed. For each database we searched three domains of exploded MeSH keyword terms. The general terms for each domain were (1) motion analysis, (2) validation and (3) laparoscopy. Where a keyword mapped to further subject headings, those considered relevant were also exploded to maximise coverage of the literature. Studies published in a foreign language were translated into English [24]. The last search date was 29 September 2011. This search strategy was undertaken by two independent reviewers, and articles retrieved according to the inclusion criteria. Articles arising from cross-referencing were also included. Duplicate articles and those

Table 1 Summary of motion analysis systems available for assessment of laparoscopic skill in general surgery

Motion analysis system	Description
Advanced Dundee Endoscopic Psychomotor Tester (ADEPT)	Consists of a dome enclosing a defined workspace that contains a target plate. Trainees are instructed to undertake up to four tasks using the target plate, including flicking a switch and turning a dial. Excessive contact with the plate or contact outside of the plate is measured as an error, recorded in seconds. Total time required to execute a task is also recorded [27].
Hiroshima University Endoscopic Surgical Assessment Device (HUESAD)	Consists of optical scale sensors, micro-encoders, an experimental table and monitor, which are connected to a computer. This enables the movement of the instrument tips to be tracked while a task is being performed. It is possible to measure two rotation angle parameters, one distance parameter and time taken [29].
Imperial College Surgical Assessment Device (ICSAD)	Utilises electromagnetic sensors placed on the dorsum of a trainee's hands, allowing hand movements to be tracked. Allows use within simulated and operating theatre environments. Data are produced by custom-built software [16].
ProMIS Augmented Reality Simulator	Trainees are able to use laparoscopic instruments to interact with the virtual environment, also including haptic feedback. Movements in space are tracked by a computer, which derives the performance metrics (time, smoothness and path length). Tasks include basic laparoscopic movements such as camera navigation as well as more complex tasks such as sharp dissection [13].
Robotic Video and Motion Analysis Software (ROVIMAS)	Translates three-dimensional coordinate data from the Isotrak II motion tracking device (Polhemus Inc., Colchester, VT) into useful motion parameters, e.g. time taken, path length and number of movements for each hand [18].

Table 2 Overview of validity types (adapted from Moorthy et al. [44])

Approach	Type of validity	Description	Method of examination
Subjective	Face validity	The extent to which the test or task resembles the real-life equivalent	Expert questionnaires
	Content validity	The extent to which the domain that is being measured is actually measured by the assessment tool	Expert questionnaires
Objective	Construct validity	The extent to which a test measures the trait that it purports to measure (for example the extent to which a test or task discriminates between various levels of expertise)	Measurement of relevant parameters between groups of variable experience
	Criterion validity		
	1. Concurrent validity	The extent to which the results of the assessment tool correlate with the gold standard for that particular domain	Comparison with patient-based data
	2. Predictive validity	The ability of the test or task to predict future performance	Correlation between test or task scores with future performance scores

clearly unrelated to the inclusion criteria were excluded. Any disagreements between the reviewers were referred to a third party.

Inclusion and exclusion criteria

All studies investigating motion analysis as a valid tool for assessment of laparoscopic skill in general surgery were included. Inclusion criteria included: sufficient detail of motion analysis technology used (including information regarding the precise motion metrics measured), description of the tasks being investigated and the type of validity measured. Studies that validated laparoscopic simulators for which motion analysis did not form the primary method of assessment were excluded. Furthermore, studies were excluded if they were validating assessment tools in specialities other than general surgery and/or if motion analysis was being validated for *laparoscopic training* rather than assessment. Evidence validating motion analysis for laparoscopic training is limited, and its inclusion would lead to further study design heterogeneity. Review articles and conference abstracts were also excluded.

Outcome measures and analysis

Each of the studies included was rated according to a modified form of the Oxford Centre for Evidence-Based Medicine (CEBM) levels of evidence and recommendation [25, 26]. Information was extracted from each study in accordance to the inclusion criteria. Common endpoints between studies were identified and compared when statistically significant results were reported, the principle summary statistic being the difference in means or medians. It was judged that the data were not suitable for meta-analysis due to study design heterogeneity.

Results

The primary search identified 2,039 records. Three hundred and eighty-eight duplicates were removed, and the remaining 1,651 abstract records screened for relevance. Following this process, 1,522 records were excluded and 129 full-text articles obtained. Full-text review excluded a further 124 studies, while cross-referencing identified 8 studies. At the end of this process, 13 studies were included for review (Fig. 1). These studies investigated four different motion analysis devices: the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT; two studies [27, 28]),

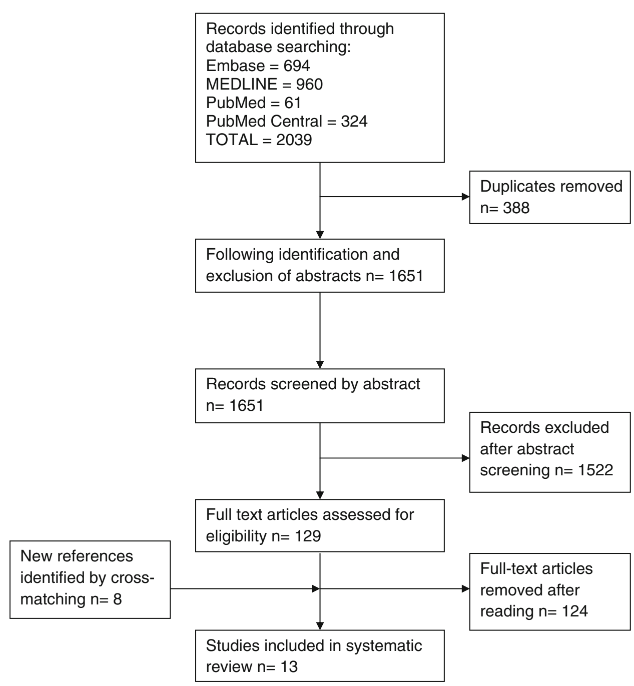

Fig. 1 PRISMA [23] flow diagram for selection of studies

Table 3 Studies included in review

Motion analysis	Authors
ADEPT	Macmillan et al. [27]
	Francis et al. [28]
HUESAD	Egi et al. [29]
	Tokunaga et al. [30]
ICSAD	Smith et al. [9]
	Moorthy et al. [31]
	Xeroulis et al. [32]
ProMIS	Van Sickle et al. [13]
	Broe et al. [33]
	Oostema et al. [34]
	Pellen et al. [35]
	Pellen et al. [36]
ROVIMAS	Aggarwal et al. [18]

the Hiroshima University Endoscopic Surgical Assessment Device (HUESAD; two studies [29, 30]), the Imperial College Surgical Assessment Device (ICSAD; three studies [9, 31, 32]), the ProMIS Augmented Reality Simulator (five studies [13, 33–36]) and the Robotic and Video Motion Analysis Software (ROVIMAS; one study [18]) (Table 3). No randomized controlled trials (RCTs) were identified. Twelve studies were graded at level 2b evidence [9, 13, 18, 28–36], and one study at level 3 [27].

Construct validity

Construct validity was examined in 12 (92.3 %) studies [9, 13, 18, 28–36]. There was a large degree of variation between studies, in terms of both group allocation and methodology (Table 4). Comparison between common endpoints (Table 5) was made in order to provide the following levels of recommendation (Table 6):

ADEPT: One study confirmed construct validity for the error score endpoint [28], when comparing novices and experts (level 3 recommendation).

HUESAD: Two studies established construct validity for the following endpoints: time taken to complete task [29, 30] (level 2 recommendation), deviation from ideal vertical and horizontal planes [29] and approaching time [30] (level 3 recommendation) when comparing novices and experts during a navigation task.

ICSAD: Three studies reported construct validity for the following endpoints: time (stage 1, 2 and 4 [9], tasks 1, 2, 3 and 4 [32]), number of hand movements (stage 1, 2 and 4 [9], task 1, 3 and 4 [32]) and path length (stage 1 and 2 [9], task 1 and 4 [32]) when comparing novices, intermediates and experts in the following tasks: laparoscopic cholecystectomy (LC) [9] and fundamentals of laparoscopic surgery (FLS) tasks [32] (all level 2 recommendations). Moorthy et al. [31] reported construct validity for time and path length in their laparoscopic suturing task when novices were compared with intermediates, and intermediates with experts. Two of the studies also demonstrated construct validity of overall expert rating scales that were used alongside motion analysis (level 2 recommendation) [31, 32].

ProMIS: Five studies established construct validity for the following endpoints: time [13, 33–36], path length [13, 34–36], smoothness of movement [13, 34–36] (level 2 recommendation) and number of hand movements [33] (level 3 recommendation) when comparing novices versus experts [13], novices versus intermediates versus experts [34, 35] or medical students/preregistration house officers (PRHOs) versus senior house officers versus surgical trainees versus consultants [33, 36] in various laparoscopic bench tasks. The tasks included suturing [13], orientation [33, 34, 36], object positioning [34, 36], knot tying [34] and sharp dissection [34–36].

ROVIMAS: One study confirmed construct validity for the following endpoints: time (overall, stage 1, 2 and 3), number of hand movements (stage 1) and path length (stage 1), when comparing novices and experts in a real-life LC [18] (level 3 recommendation). Number of hand movements and path length were both unable to distinguish between novices and experts in clipping and cutting the cystic duct (stage 2) and artery (stage 3), or during dissection (stage 4) [18].

Other validity types

Face validity was reported by one study for ADEPT [27] and one study for ICSAD [9] (no data provided). Three studies reported concurrent validity [9, 27, 35]. Macmillan et al. state that for ADEPT a high correlation was seen between the number of perfect runs and blinded clinical assessments (Spearman's rho 0.74) [27]. Concurrent validity was also confirmed by one study for ICSAD [31], and ProMIS [35], through the observation that motion analysis metrics correlated with expert and global rating scores (ICSAD: path length, Spearman's rho –0.78, $p = 0.000$; ProMIS: time and path length, Spearman's rho 0.88, $p < 0.05$) (all level 3 recommendations). None of the 13 studies included in this systematic review investigated content or predictive validity.

Discussion

This study presents the evidence for the use of motion analysis in laparoscopic skills assessment. A previous review by van Hove et al. [15] assessed a range of objective tools available to assess surgical skill, including

Table 4 Summary of methods

Motion analysis	Reference	*n*	Groups	Validation	Task	Endpoints
ADEPT	Macmillan et al. [27]	10	10 HSTs	Predictive and face	10 repeats of ADEPT tasks (Table 1)	Execution time, plate error score and probe error score
	Francis et al. [28]	40	$N = 20$ HSTs $E = 20$ consultants	Construct	ADEPT tasks (Table 1)	Execution time, plate error score, probe error score and overall performance
HUESAD	Egi et al. [29]	37	$N = 25$ medical students (no exp) $E = 12$ surgeons (>100 lap procedures)	Construct	Navigation	Time and deviation from the ideal course in the vertical and horizontal planes
	Tokunaga et al. [30]	36	$N = 20$ medical students $E = 16$ (>50 lap procedures)	Construct	Navigation	Total time, approaching time and intermediate time*
ICSAD	Smith et al. [9]	15	$N = 5$ (<10 human LCs) $I = 7$ (10–100 human LCs) $E = 3$ (>100 human LCs)	Construct and face	LC (porcine model) subdivided into four stages	Time, distance travelled, speed of movement and number of hand movements
	Moorthy et al. [31]	26	$N = 13$ (<10 LCs, no lap suturing) $I = 7$ (10–50 LCs, <50 lap sutures) $E = 6$ (>100 lap suturing procedures)	Construct and concurrent	Laparoscopic suturing (laboratory based)	Time and distance travelled
	Xeroulis et al. [32]	26	$N = 13$ (PGY 1–3) $I = 7$ (PGY 4–5) $E = 6$ staff surgeons	Construct	FLS education modules (four tasks)***	Time, distance travelled, number of hand movements and expert rating scores
ProMIS	Van Sickle et al. [13]	10	$N = 5$ (medical students) $E = 5$ (HSTs, significant lap exp)	Construct	Laparoscopic suturing	Time, path length, smoothness of movement‡, and error score
	Broe et al. [33]	20	Group 1 = 7 PRHOs Group 2 = 6 SHOs Group 3 = 1 JSpRs Group 4 = 3 HSTs Group 5 = 3 consultants	Construct	Laparoscopic orientation	Time, number of movements and OSATS global scoring
	Oostema et al. [34]	47	$N = 24$ medical students $I = 19$ PGY 1–5 $E = 3$ consultants	Construct	Laparoscopic orientation, object positioning, sharp dissection and knot tying	Time, path length and smoothness of movement
	Pellen et al. [35]	30	$N = 10$ medical students (no lap exp) $I = 10$ HSTs (<50 lap procedures) $E = 10$ consultants (>100 lap procedures)	Construct and concurrent	Sharp dissection (laboratory based)	Time, path length, smoothness and procedure-specific rating scale
	Pellen et al. [36]	160	Group 1 = 53 medical students Group 2 = 28 BSTs Group 3 = 61 HSTs Group 4 = 18 consultants (>100 lap procedures)	Construct	Laparoscopic orientation, object positioning and sharp dissection	Time, path length and smoothness

Table 4 continued

Motion analysis	Reference	n	Groups	Validation	Task	Endpoints
ROVIMAS	Aggarwal et al. [18]	19	N = 6 (<10 human LCs) E = 13 (>100 human LCs)	Construct	LC (on patient)	Time, distance travelled and number of hand movements

BST basic surgical trainee; *HST* higher surgical trainee; *JSpR* junior specialist registrar; *SSpR* senior specialist registrar; *N* novice; *I* intermediate; *E* expert; *LC* laparoscopic cholecystectomy; *PGY* postgraduate year; *PRHO* preregistration house officer; *SHO* senior house officer; *OSATS* Objective Structured Assessment of Technical Skills

* Approaching time = time taken to move between two points in HUESAD navigation tasks. Intermediate time = total time − approaching time

** Information regarding grade/experience not available

*** Fundamentals of laparoscopic surgery (FLS) is a CD-ROM-based education module for hands-on skills-based training [46]

‡ Smoothness is defined as the number of times an instrument changes velocity during completion of a task [34]

motion analysis. However, this did not provide information regarding the precise surgical skill assessed, nor did it provide subsequent levels of recommendation. The authors included studies validating the TrEndo Tracking System, which so far has only been studied in obstetrics and gynaecology trainees [37, 38]. These studies have produced promising results, and we recommend further studies investigating its application within general surgery. Carter et al. [26] published consensus guidelines concerning evidence rating and subsequent levels of recommendation for evaluation and implementation of simulators and skills training programmes [25, 26]. The authors produced an alternative system due to the absence of published validation studies that have rigorous experimental methodology [26]. Our review utilises this version of the CEBM system, and actual levels of recommendation for each tool have been provided for the first time (Table 6).

This review reports construct validity for a range of different motion analysis metrics across three different training environments (VR [13, 33–36], laboratory based [9, 28–32] and the operating theatre [18]). The most commonly validated metrics were time to complete a task, path length and number of hand movements. One ICSAD study attempted to establish construct validity for velocity during a simulated porcine LC model [9]. Velocity is a function of time and path length, both of which were also measured. However, while velocity was found to largely lack construct validity, this was not the case for time and path length. Smith et al. explain this by stating that each movement made by experienced surgeons is more efficient, meaning that, while the speed of movements is not significantly quicker, instead they are more goal directed so that tasks are completed in less time [9]. Despite only being a metric measured by the ProMIS simulator, smoothness of movements was also consistently shown to discriminate between different levels of experience [13, 33–36].

Aggarwal et al. [39] state the importance of breaking down training and assessment into basic, intermediate and advanced stages. It could be suggested that ADEPT and HUESAD could be used to assess basic training as they utilise simple orientation and movement skills in a non-anatomical environment. ICSAD, ProMIS and ROVIMAS could be used to assess intermediate competence. There are animal tissue models and virtual reality simulators that exist for a range of general surgery procedures that could be used in conjunction with these motion analysis technologies. This has already been demonstrated in a porcine model for LC [9], and adaptations to the devices may enable their use in endoscopy training. The flexibility of use offered by ICSAD and ROVIMAS means that advanced competency could be assessed. Construct validity during a real-life LC has already been demonstrated for ROVIMAS [18].

This systematic review also showed that very few forms of validity are being examined apart from construct. The more forms of validity that are demonstrated, the stronger the overall argument for the use of a particular technology [20]. While two studies report face validity [9, 27], no expert rating data were provided to support this. It may not be possible to face-validate motion analysis technology, as any attempt to do so would be assessing the realism of the laparoscopic set-up instead. While it is important to establish construct validity for each endpoint and in every procedure that motion analysis may eventually be used to assess, its practical use in real-life assessment is limited. Predictive validity represents a more useful modality to investigate, and it is unfortunate that there have been no studies undertaken to investigate this.

The main limitation of this review is the degree of methodological variation between included studies, which prevented meta-analysis. The largest degree of variation was seen for group allocation, which was largely based on career grades, although most studies used further inclusion criteria within each grade based on varying levels of laparoscopic experience. This limitation is explained by the fact that number of procedures performed is a

Table 5 Summary of statistical outcomes for construct validity studies

Motion analysis	Reference	Group/task	Time	Velocity	Path length*	No. of hand movements	Smoothness	↕ Deviation	↔ Deviation	Error score	Approaching time	Int. time	Global rating
ADEPT	Francis et al. [28]	*N* vs. *E*	0.420	–	–	–	–	–	–	**0.007**	–	–	0.400
HUESAD	Egi et al. [29]	*N* vs. *E*: task 1	**<0.0001**	–	–	–	–	**0.009**	**0.0004**	–	–	–	–
		N vs. *E*: task 2	**<0.0001**	–	–	–	–	**0.0002**	**<0.0001**	–	–	–	–
	Tokunaga et al. [30]	*N* vs. *E*: task 1	**<0.0001**	–	–	–	–	–	–	–	**<0.0001**	>0.050	–
		N vs. *E*: task 2	**<0.0001**	–	–	–	–	–	–	–	**<0.0001**	>0.050	–
ICSAD	Smith et al. [9]	*N* vs. *I* vs. E: Calot's triangle	**<0.000**	NS	**0.0120**	**<0.000**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: clipping duct/artery	**<0.000**	NS	**0.002**	**<0.000**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: cutting duct/artery	NS	NS	NS	NS	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: dissection	**<0.000**	**0.032**	NS	**0.049**	–	–	–	–	–	–	–
	Moorthy et al. [31]	*N* vs. *I* vs. *E*	**0.000**	–	**0.000**	–	–	–	–	–	–	–	**0.000**
		N vs. *I*	**0.060**	–	**0.010**	–	–	–	–	–	–	–	–
		I vs. *E*	**0.001**	–	**0.000**	–	–	–	–	–	–	–	–
	Xeroulis et al. [32]	*N* vs. *I* vs. *E*: overall	–	–	–	–	–	–	–	–	–	–	**0.001**
		N vs. *I* vs. *E*: task 1	**0.001**	–	**0.025**	**0.050**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: task 2	**0.002**	–	0.286	0.297	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: task 3	**0.005**	–	0.121	**0.011**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: task 4	**0.001**	–	**0.001**	**0.002**	–	–	–	–	–	–	–
ProMIS	Van Sickle et al. [13]	*N* vs. *E*	**<0.0001**	–	**<0.0001**	–	**<0.0001**	–	–	–	–	–	
	Broe et al. [33]	1 vs. 4 and 5	**<0.050**	–	–	**<0.050**	–	–	–	–	–	–	–
	Oostema et al. [34]	*N* vs. *I* vs. *E*: camera navigation	**0.09**	–	0.900	–	**0.030**	–	–	–	–	–	–
		N vs. *I* vs. *E*: object positioning	**<0.010**	–	**0.010**	–	**<0.010**	–	–	–	–	–	–
		N vs. *I* vs. *E*: sharp dissection	**<0.010**	–	**<0.010**	–	**<0.010**	–	–	–	–	–	–
		N vs. *I* vs. *E*: knot tying	**<0.010**	–	**<0.010**	–	**<0.010**	–	–	–	–	–	–
	Pellen et al. [35]	*N* vs. *I* vs. *E*	**0.010**	–	**0.010**	–	**0.010**	–	–	0.143	–	–	**<0.050**
	Pellen et al. [36]	1 vs. 2 vs. 3 vs. 4 vs. 5: orientation	0.058	–	0.097	–	**0.008**	–	–	–	–	–	–
		1 vs. 2 vs. 3 vs. 4 vs. 5: object positioning	**0.000**	–	**0.000**	–	**0.000**	–	–	–	–	–	–
		1 vs. 2 vs. 3 vs. 4 vs. 5: sharp dissection	**0.000**	–	**0.000**	–	**0.000**	–	–	**0.001**	–	–	–
ROVIMAS	Aggarwal et al. [18]	*N* vs. *E*: overall	**0.036**	–	0.625	0.389	–	–	–	–	–	–	–
		N vs. *E*: Calot's triangle	**0.002**	–	**0.048**	**0.007**	–	–	–	–	–	–	–
		N vs. *E*: clip and cut duct	**0.013**	–	0.063	0.553	–	–	–	–	–	–	–
		N vs. *E*: clip and cut artery	**0.002**	–	0.119	0.204	–	–	–	–	–	–	–
		N vs. *E*: dissection	0.471	–	0.377	0.396	–	–	–	–	–	–	–

NS no significant difference (where *p* values not provided); *N* novice; *I* intermediate; *E* expert

– Denotes parameter not measured in study/data not provided

Bold indicates *p* value significant at 0.05 level

* Equivalent to distance moved by instrument

Table 6 Level of evidence and recommendation for each motion analysis device

Motion analysis device	Type of validity (level of evidence)			Recommendation level
	Face	Concurrent	Construct	
ADEPT	Yes (3) [27]§	Yes (3) [27]	Yes, for error score (2b) [28]	Level 4 for face and concurrent validity Level 3 for construct validity endpoints
HUESAD	No	No	Yes, for time taken to complete task (2b) [29, 30] (*) Yes, for deviation from ideal vertical plane (2b) [29] Yes, for deviation from ideal horizontal plane (2b) [29] (**) Yes, for approaching time (2b) [30]	Level 2 for (*) construct validity endpoints; level 3 for (**) construct validity endpoints.
ICSAD	Yes (2b) [9]§	Yes (2b) [31]	Yes, for time taken to complete task (2b) [9, 31, 32] Yes, for number of hand movements (2b) [9, 31, 32] Yes, for path length (2b) [9, 31, 32]	Level 3 for face validity All level 2 for construct validity endpoints
ProMIS	No	Yes (2b) [35]	Yes, for time taken to complete task (2b) [13, 33–36] Yes, for path length (2b) [13, 34–36] (*) Yes, for smoothness (2b) [13, 34–36] Yes, for number of hand movements (2b) [33] (**)	Level 3 for concurrent validity Level 2 for (*) construct validity endpoints; level 3 for (**) construct validity endpoints
ROVIMAS	No	No	Yes, for time taken to complete task (2b) [18] Yes, for number of hand movements (2b) [18] Yes, for path length (2b) [18]	All level 3 for construct validity endpoints

§ No expert rating data provided; Macmillan et al. [27] state face validity is assured as the study utilised the same equipment used in minimal access surgery. Smith et al. [9] state face validity is assured due to the observation that there was little change in performance amongst members of the same group

non-objective measure of experience. A more objective approach to group allocation could have been made on the basis of Objective Structured Assessment of Technical Skills (OSATS) scoring. A further limitation is that the majority of the studies included compared groups across wide ranges of experience (e.g. novice versus intermediate versus expert), where outcomes may be largely dependent on the novice versus expert element of this analysis. Motion analysis must demonstrate the sensitivity to discriminate between all individual grades if it is to be used to assess laparoscopic competence.

Motion analysis does carry some limitations which require discussion. Firstly, many of the devices require calibration to account for individual physiological tremor. This may require technical support during each procedure. Additionally, there is the issue of cost, which may prevent widespread use across all training centres.

In order for motion analysis to be used as an assessment tool it must be shown to work in a real-life environment. While the feasibility of using motion analysis in a real-life operating theatre has been demonstrated for ICSAD [40] and ROVIMAS [18], the correlation between motion analysis assessment in the laboratory and its subsequent use within the operating theatre needs to be evaluated. Quantitative assessment outcomes must be shown to be equivalent between different training environments, otherwise the application of motion analysis to provide trainee feedback is undermined.

Using motion analysis in isolation may remove the user from the context of the operating theatre. As surgical competence is multimodal, it is important that assessment is not only based on specific outcomes (such as dexterity) but also global outcomes, such as task accuracy and outcome. This is made possible through the dual application of motion analysis alongside global checklists [e.g. Global Operative Assessment of Laparoscopic Skills (GOALS) and Objective Structured Assessment of Technical Skills (OSATS)] [41]. Furthermore, procedure-specific rating scales have also been developed to assess specific technical aspects of different operations, including LC [42] and Nissen fundoplication [43]. Using these systems, assessment can either occur "live", whilst a trainee is undertaking a specific task [44]. Several studies included in this review included global rating scores, which were found to correlate with motion analysis metrics [18, 31, 35].

It has been suggested that surgery is 75 % decision-making and 25 % dexterity [45]. While motion analysis may provide a promising tool to assess dexterity, it cannot provide information on the numerous attributes that contribute to the other three-quarters of a good surgeon's skill set. Further work is needed to correlate motion analysis against similarly validated measurements of surgical decision-making in different scenarios.

Conclusions

We have demonstrated that there is evidence validating the use of motion analysis to assess laparoscopic skill. The most valid metrics appear to be time, path length and number of hand movements. More work is needed to establish predictive validity for each of these metrics. Future work should concentrate on the conversion of motion analysis data into competency-based scores or indices for trainee feedback.

Acknowledgments J. Ansell is currently funded by a Royal College of Surgeons England research fellowship grant.

Disclosures Authors John D. Mason, James Ansell, Neil Warren and Jared Torkington have no conflicts of interest or financial ties to disclose.

References

1. Darzi A, Smith S, Taffinder N (1999) Assessing operative skill. Using these systems, assessment. BMJ 318:887–888
2. Chesser S, Bowman K, Phillips H (2002) The European Working Time Directive and the training of surgeons. BMJ 325:S69
3. Donohoe C, Sayana M, Kennedy M, Niall D (2010) European Time Directive: implications for surgical training. Ir Med J 103:57–59
4. Smith R (1998) All changed, changed utterly: British medicine will be transformed by the Bristol case. BMJ 316:1917–1918
5. Bridges M, Diamond D (1999) The financial impact of teaching surgical residents in the operating room. Am J Surg 177:28–32
6. Munz Y, Kumar B, Moorthy K, Bann S, Darzi A (2004) Laparoscopic virtual reality and box trainers. Is one superior to the other? Surg Endosc 18:485–494
7. Sturm L, Windsor J, Cosman P, Cregan P, Hewett P, Maddern G (2008) A systematic review of skills transfer after surgical simulation training. Ann Surg 248(2):166–179
8. Chung J, Sackier J (1998) A method of objectively evaluating improvements in laparoscopic skills. Surg Endosc 12:1117–1120
9. Smith S, Torkington J, Brown T, Taffinder N, Darzi A (2002) Motion analysis. A tool for assessing laparoscopic dexterity in the performance of a laboratory-based laparoscopic cholecystectomy. Surg Endosc 16:640–645
10. Clayden G (1994) Teaching laparoscopic surgery. Preliminary training on animals is essential. BMJ 309:342
11. Wilson M, Middlebrooke A, Sutton C, Stone R, McCloy R (1997) MIST-VR: a virtual reality trainer for laparoscopic surgery assesses performance. Ann R Coll Surg Engl 79:403–404
12. Larson A (2001) An open and flexible framework for computer aided surgical training. Stud Health Technol Inform 81:263–265
13. Van Sickle K, McClusky DI, Gallagher A, Smith C (2005) Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surg Endosc 19:1227–1231
14. McDougall E, Corica F, Boker J, Sala L, Stoliar G, Borin J, Chu F, Clayman R (2006) Construct validity testing of a laparoscopic surgical simulator. J Am Coll Surg 202(5):779–787
15. van Hove P, Tuijthof G, Verdaasdonk E, Stassen LP, Dankelman J (2010) Objective assessment of technical surgical skills. Br J Surg 97:972–987
16. Datta V, Mackay S, Darzi A, Gillies D (2001) Motion analysis in the assessment of surgical skill. Comput Methods Med Biomed Eng 4:515–523
17. Aggarwal R, Hance J, Darzi A (2004) Surgical education and training in the new millenium. Surg Endosc 18:1409–1410
18. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, Bello F, Darzi A (2007) An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg 245(6):992–999
19. Thijssen A, Schijven M (2010) Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? Am J Surg 199:529–541
20. Sedlack R (2011) Validation process for new endoscopy teaching tools. Tech Gastro Endosc 13:151–154
21. Feldman L, Sherman V, Fried G (2004) Using simulators to assess laparoscopic competence: ready for widespread use? Surgery 135:28–42
22. Technical skills education surgery (2011). http://www.facs.org/education/technicalskills/faqs/faqs.html. Accessed 8 Oct 2011
23. The PRISMA (2011). http://www.prisma-statement.org/index.htm. Accessed 1 Oct 2011
24. Google Translate (2011). http://translate.google.com/. Accessed 30 Sept 2011
25. Oxford Centre for Evidence Based Medicine—Levels of Evidence (2011). http://www.cebm.net/index.aspx?o=1025. Accessed 30 Sept 2011
26. Carter F, Schijven M, Aggarwal R, Grantcharov T, Francis N, Hanna G, Jakimowicz J (2005) Work group for evaluation and implementation of simulators and skills training programmes. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 19:1523–1532
27. Macmillan A, Cuschieri A (1999) Assessment of innate ability and skills for endoscopic manipulations by the advanced Dundee endoscopic psychomotor tester: predictive and concurrent validity. Am J Surg 177:274–277
28. Francis N, Hanna G, Cuschieri A (2002) The performance of master surgeons on the advanced Dundee endoscopic psychomotor tester. Arch Surg 137:841–844
29. Egi H, Okajima M, Yoshimitsu M, Ikeda S, Miyata Y, Masugami H, Kawahara T, Kurita Y, Kaneko M, Asahara T (2008) Objective assessment of endoscopic surgical skills by analyzing direction-dependent dexterity using the Hiroshima University endoscopic surgical assessment device (HUESAD). Surg Today 38:705–710
30. Tokunaga M, Egi H, Minoru H, Yoshimitsu M, Sumitani D, Kawahara T, Okajima M, Ohdan H (2011) Approaching time is important for assessment of endoscopic surgical skills. Min Invasive Ther Allied Technol 21:142–149
31. Moorthy K, Munz Y, Doris A, Bello F, Chang A, Darzi A (2004) Bimodal assessment of laparoscopic suturing skills: construct and concurrent validity. Surg Endosc 18:1608–1612
32. Xeroulis G, Dubrowski A, Leslie K (2009) Simulation in laparoscopic surgery: a concurrent validity study for FLS. Surg Endosc 23:161–165

33. Broe D, Ridgway P, Johnson S, Tierney K, Conlon K (2006) Construct validation of a novel hybrid surgical simulator. Surg Endosc 20:900–904
34. Oostema J, Abdel M, Gould J (2008) Time-efficient laparoscopic skills assessment using an augmented-reality simulator. Surg Endosc 22:2621–2624
35. Pellen M, Horgan L, Barton J, Attwood S (2009) Laparoscopic surgical skills assessment: can simulators replace experts? World J Surg 33:440–447
36. Pellen M, Horgan L, Barton J, Attwood S (2009) Construct validity of the ProMIS laparoscopic simulator. Surg Endosc 23:130–139
37. Chmarra M, Kolkman W, Jansen F, Grimbergen C, Dankelman J (2007) The influence of experience and camera holding on laparoscopic instrument movements measured with the TrEndo tracking system. Surg Endosc 21:2069–2075
38. Chmarra M, Klein S, de Winter J, Jansen F-W, Dankelman J (2010) Objective classification of residents based on their psychomotor laparoscopic skills. Surg Endosc 24:1031–1039
39. Aggarwal R, Moorthy K, Darzi A (2004) Laparoscopic skills training and assessment. Br J Surg 91:1549–1558
40. Doris A, Aggarwal R, Bello F, Moorthy K, Munz Y, Gillies D, Darzi A (2005) Synchronized video and motion analysis for the assessment of procedures in the operating theater. Arch Surg 140:293–299
41. Martin J, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchinson C, Brown M (1997) Objective Structured Assessment of Technical Skill (OSATS) for surgical residents. Br J Surg 84:273–278
42. Eubanks T, Clements R, Pohl D, Williams N, Schaad D, Horgan S, Pellegrini C (1999) An objective scoring system for laparoscopic cholecystectomy. J Am Coll Surg 189:566–574
43. Dath D, Regehr G, Birch D, Schlachta C, Poulin E, Mamazza J, Reznick R, MacRae H (2004) Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. Surg Endosc 18(12):1800–1804
44. Moorthy K, Munz Y, Sarker S, Darzi A (2003) Objective assessment of technical skills in surgery. BMJ 327:1032–1037
45. Spencer F (1978) Teaching and measuring surgical techniques—the technical evaluation of competence. Bull Am Coll Surg 63:9–12
46. FLS program (2011). http://www.flsprogram.org. Accessed 9 Oct 2011