

An Evaluation of the Feasibility, Validity, and Reliability of Laparoscopic Skills Assessment in the Operating Room

Rajesh Aggarwal, MRCS,* Teodor Grantcharov, PhD,† Krishna Moorthy, MD,* Thor Milland, MD,‡ Pavlos Pappasavas, MD,‡ Aristotelis Dosis, PhD,* Fernando Bello, PhD,* and Ara Darzi, MD*

Objective: To assess the use of a synchronized video-based motion tracking device for objective, instant, and automated assessment of laparoscopic skill in the operating room.

Summary Background Data: The assessment of technical skills is fundamental to recognition of proficient surgical practice. It is necessary to demonstrate the validity, reliability, and feasibility of any tool to be applied for objective measurement of performance.

Methods: Nineteen subjects, divided into 13 experienced (performed >100 laparoscopic cholecystectomies) and 6 inexperienced (performed <10 LCs) surgeons completed LCs on 53 patients who all had a diagnosis of biliary colic. Each procedure was recorded with the ROVIMAS video-based motion tracking device to provide an objective measure of the surgeon's dexterity. Each video was also rated by 2 experienced observers on a previously validated operative assessment scale.

Results: There were significant differences for motion tracking parameters between the 2 groups of surgeons for the Calot triangle dissection part of procedure for time taken ($P = 0.002$), total path length ($P = 0.026$), and number of movements ($P = 0.005$). Both motion tracking and video-based assessment displayed intertest reliability, and there were good correlations between the 2 modes of assessment ($r = 0.4$ to 0.7 , $P < 0.01$).

Conclusions: An instant, objective, valid, and reliable mode of assessment of laparoscopic performance in the operating room has been defined. This may serve to reduce the time taken for technical skills assessment, and subsequently lead to accurate and efficient audit and credentialing of surgeons for independent practice.

(*Ann Surg* 2007;245: 992–999)

Recent publications of the rates of medical errors and adverse events within health care, and particularly during surgery, have drawn the spotlight toward the meth-

ods of credentialing surgeons to perform procedures independently.^{1–4} Training boards and certifying bodies are coming under increasing pressure to ensure individuals demonstrate the necessary skills to perform operations safely.^{5–9} This is not only important for patient safety but also underpins the development of a proficiency-based training curriculum.

It is somewhat surprising then that there are no tools in widespread use that are feasible, valid, and reliable for assessment of technical surgical skill (Table 1).¹⁰ Current training outcomes are assessed by live evaluations of the trainee by the master surgeon, a process that is known to be biased and subjective.¹¹ More objective data are available from morbidity and mortality data, although this is rarely a sole function of operative skill and thus does not truly reflect an individual's surgical competence.¹² The majority of trainees also maintain a log of the procedures performed, but these are indicative merely of procedural performance rather than a measure of technical ability.¹³

Although a number of new tools have been developed to assess surgical technical performance, their use remains within the confines of surgical skills laboratories.¹⁰ These include virtual reality simulators and psychomotor training devices, which are designed primarily to assess performance during critical parts of a procedure, rather than a complete operation.¹⁴ The realism (or face validity) of such simulations is not perfect and the situations lack context, leading to a failure of operators to treat the models like real patients.¹⁵

The ideal device for objective assessment of real surgical procedures would be one that can automatically, instantly, and objectively provide feasible, valid, and reliable data regarding performance within the operating room.¹⁶ It is with this approach that our Department has developed the ROVIMAS motion tracking software, which enables surgical dexterity to be quantified and thus reported instantly by a computer program.¹⁷ Although automatic, objective, and instant, the data do not provide any information regarding the quality of the procedure performed. The system does, however, incorporate the ability to synchronously record video of the operative procedure, which can then be evaluated according to a valid and reliable rating scale. This can enable a definition of dexterity not only for whole procedures, but also for critical steps of a particular procedure.

A preliminary publication has confirmed the feasibility of using the device within the operating room to assess laparoscopic skills.¹⁷ The primary aim of this study was to

From the *Department of Biosurgery & Surgical Technology, Imperial College London, UK; †Department of Surgical Gastroenterology, Glostrup University Hospital, Glostrup, Denmark; and ‡Department of Surgery, Western Pennsylvania Hospital, Pittsburgh, PA.

Reprints: Rajesh Aggarwal, MRCS, Department of Biosurgery & Surgical Technology, Imperial College London, 10th Floor, QEQM Building, St. Mary's Hospital, Praed Street, London, W2 1NY. E-mail: rajesh.aggarwal@imperial.ac.uk.

Copyright © 2007 by Lippincott Williams & Wilkins
ISSN: 0003-4932/07/24506-0992
DOI: 10.1097/01.sla.0000262780.17950.e5

TABLE 1. Qualities of the Ideal Surgical Assessment Tool

Term	Definition
Feasibility	Measure of whether something is capable of being done or carried out
Validity	
Face validity	Extent to which the examination resembles real life situations
Content validity	Extent to which the domain that is being measured is measured by the assessment tool—for example, while trying to assess technical skills we may actually be testing knowledge
Construct validity	Extent to which a test measures the trait that it purports to measure; one inference of construct validity is the extent to which a test discriminates between various levels of expertise
Concurrent validity	Extent to which the results of the assessment tool correlate with the gold standard for that domain
Predictive validity	Ability of the examination to predict future performance
Reliability	
Test-retest	Measure of a test to generate similar results when applied at two different points
Inter-rater	Measure of the extent of agreement between two or more observers when rating the performance of an individual

Adapted from Moorthy et al.¹⁰

determine the validity and reliability of a new concept for technical skills assessment in the operating room, a combination of motion analysis and video assessment. Both approaches have been individually validated in the literature, although the introduction of a hybrid between the 2 modes of assessment has not been previously attempted.

METHODS

Subjects

Nineteen surgeons were recruited to the study, and subdivided into 6 novice (<10 laparoscopic cholecystectomies, LCs) and 13 experienced (>100 LCs) practitioners. The aim was for each surgeon to perform a minimum of 2 procedures, with consecutive cases recorded over a period of 6 months.

Operative Procedure

LC was chosen as the operative procedure of choice as it is a common operation, performed in a fairly standardized manner and amenable to both motion tracking and video-based analysis. Furthermore, LC is an index procedure for commencement of training and ongoing assessment of laparoscopic skills.^{5,9}

Patients

Ethical approval was obtained from the local research ethics committee to record video data of each operation. Patients were recruited from 2 surgical departments, and all were consented prior to entry into the study. To reduce the

effect of disease and patient variability, all patients recruited to the study were deemed to have a diagnosis of biliary colic. To enable objectification of this approach, inclusion and exclusion criteria were classified according to patient and disease state, and upon posthoc review of the video tape according to operative state (Table 2).¹⁸

Motion Analysis Device

All procedures were recorded with the ROVIMAS software. The parameters used for this study were those that have been previously validated on bench-top assessments of laparoscopic skill on a porcine model, ie, time taken, path length, and number of movements for each hand.¹⁹

The purpose of the study was explained to all surgeons prior to the patient consent process. Once scrubbed, surgeons wore one pair of sterile gloves over their hands. Sensors were then placed onto the dorsum of each hand, followed by the donning of surgical gown and a further pair of sterile gloves. This avoided the need to sterilize the electromagnetic sensors, and friction between the gloves maintained the sensors in the correct position. Once the patient had been anesthetized, the electromagnetic emitting device was placed onto their sternum, fixed firmly by Micropore tape (3M Corporation, St. Paul, MN).

Video-Based Assessment

The video feed from the laparoscopic stack was recorded onto the ROVIMAS software through a digital video link (I-link, IEEE-1394) to a laptop computer. Recording commenced upon entry of the endoscopic camera into the peritoneal cavity and was complete upon removal of the camera from the abdomen for the final time. The open parts of the procedure were not recorded, the aim being to solely assess the laparoscopic skills of the subjects. Complete, unedited videos of each procedure were recorded with the software into Microsoft Windows .avi format (Microsoft Corporation, Redmond, WA). All data files were coded by an alphanumeric code to ensure the identity of the operating surgeon and patient were blinded to the reviewers.

The objective structured assessment of technical skill (OSATS) proposes a generic evaluation of surgical performance through use of a global rating scale.²⁰ The scale was initially validated through live-marking of bench-top tasks and is said to “boast high reliability and show evidence of validity.”²¹ The aim was to determine the validity, inter-rater, and intertest reliability of this scale for assessment of laparoscopic technical skills. Rating on the scale was performed by 2 experienced laparoscopic surgeons (T.G. and K.M.) who were blinded as to the identities of the operating surgeons.

Data Collection and Analysis

Three-dimensional coordinate data from the Isotrak II motion tracking device (Polhemus Inc, Colchester, VT) were translated into useful parameters of time taken, path length, and number of movements of each hand by the ROVIMAS software (see Dosis et al for a detailed description¹⁷). With the aid of the synchronization feature of ROVIMAS, data were derived for the entire procedure, and also for predefined

TABLE 2. Inclusion and Exclusion Criteria for Patient Entry to the Study

Characteristic	Inclusion Criteria	Exclusion Criteria
Patient characteristics		
Age	>18 yr and <65 yr	<18 yr and >65 yr
Obesity	BMI <30 kg/m ²	BMI >30 kg/m ²
Anaesthetic risk	ASA = 1 or 2	ASA >2
Hospital admission with gallbladder pathology	No	Yes
Disease characteristics		
Diagnosis	Biliary colic	Acute cholecystitis
Complications of gallstones	None	Any
ERCP	No	Yes
Blood tests (at any time preoperatively)	WCC <11 CRP <5 LFTs, normal range	WCC >11 CRP >5 LFTs abnormal
Ultrasound findings (at any time preoperatively)	Gallstones/sludge	Thickened gallbladder wall Pericholecystic fluid Ultrasonographic Murphy positive Common bile duct stone Common bile duct dilatation
Intraoperative characteristics (modified from Hanna et al.)		
Degree of difficulty	Cystic duct seen on retraction of gallbladder Unobstructed view of Calot's triangle, or Fat over Calot's triangle No obvious ductal or vascular anomaly None or filmy/loose areolar adhesions to gallbladder	Contracted, inflamed, or densely adherent gallbladder Gallbladder neck adherent to bile duct Fat-laden falciform Hypertrophied liver: quadrate lobe partially obstructing view and/or right hepatic lobe making retraction difficult Difficult, obscure, abnormal anatomy Dense omental adhesions to gallbladder Duodenal adhesions to gallbladder Stone impacted in neck or Hartmann's pouch

parts of the procedure, classified as in Table 3. It must be noted that the values for the whole procedure are not a sum of all the parts identified, ie, insertion of accessory ports, division of adhesions, removal of gallbladder, etc.

Power analysis was based upon results from a previous study on motion analysis of porcine LCs, and revealed a sample size of 20 cases per group.¹⁹ Statistical analysis used nonparametric tests of significance. Construct validity was determined by comparison of performance between novice and experienced surgical groups for dexterity parameters and scores from video-rating scales with the Mann-Whitney *U* test. Cronbach's alpha test statistic was used to ascertain the inter-rater reliability of the video-based scoring system. Intertest reliability was assessed by comparison of the first and second consecutive procedure performed by each surgeon, once again with Cronbach's alpha test statistic.

To investigate the existence of a relationship between dexterity analysis and the video-based rating scale for assessment of surgical performance, correlations between the 2 methods were calculated with the nonparametric Spearman's rank correlation coefficient. For all tests, $P < 0.05$ was considered statistically significant.

RESULTS

Procedures Performed

A total of 53 procedures were performed by the 19 surgeons recruited to the study. Six cases were excluded in an independent manner by both reviewers upon the basis of intraoperative characteristics (Table 2). Of the remaining 47 cases, 14 were performed by 6 novice surgeons and 33 by the 13 experienced surgeons. The median number of cases carried out by each surgeon was 2 (range, 1–5 cases).

Motion Tracking Data

A comparison between LCs performed by novice and experienced surgeons revealed significant differences in time taken for the whole procedure (median 2175 vs. 1979 seconds, $P = 0.036$), although not for total path length or number of movements (Table 4). This result was replicated for "clip and cut duct" (55 vs. 33 seconds, $P = 0.013$) and "clip and cut artery" (37 vs. 21 seconds, $P = 0.004$). Only dissection of Calot triangle produced significant differences between the performance of novice and experienced surgeons for all 3 motion tracking parameters (Figs. 1–3): time taken (854 vs. 393 seconds, $P = 0.002$), total path length (138 vs.

TABLE 3. Definitions of the Tasks of a Laparoscopic Cholecystectomy

	Start	Finish
Whole procedure	First moment of insertion of endoscopic camera	Final removal of endoscopic camera
Dissection of Calot's triangle	First moment gallbladder is grasped at Calot's triangle	Entry of clip applicator to the operative field of view
Clip and cut cystic duct	First entry of clip applicator to the operative field of view prior to clipping the cystic duct	Cystic duct is clipped and divided
Clip and cut cystic artery	First entry of clip applicator to the operative field of view prior to clipping the cystic artery	Cystic artery is clipped and divided
Dissection of gallbladder from liver bed	Following division of duct and artery, the first moment that the peritoneum between gallbladder and liver bed is grasped	Gallbladder is freed from liver

73 m, $P = 0.026$) and total number of movements (640 vs. 367, $P = 0.005$).

Fifteen of the 19 surgeons performed 2 or more procedures each. The intertest reliability between their first 2 consecutive cases for time taken to perform the whole procedure was $\alpha = 0.502$. With regard to dissection of Calot triangle, intertest reliability was calculated for time taken ($\alpha = 0.623$), total path length ($\alpha = 0.229$), and total number of movements ($\alpha = 0.522$).

Video-Based Data

The generic OSATS global rating scale demonstrated a significant difference in scores between the novice and experienced surgeons (median 24 vs. 27, $P = 0.031$), with an inter-rater reliability coefficient of $\alpha = 0.72$ (Fig. 4). The intertest reliabilities of the 15 surgeons who performed 2 or more procedures for the OSATS was $\alpha = 0.72$.

Comparison of Motion Tracking and Video-Rating Scales

The correlations between scores obtained from the OSATS global rating scale and validated motion tracking

parameters are shown in Table 5. All r values were statistically significant and ranged from between 0.4 to 0.7, indicating that there were good correlations between the 2 modes of assessment.

DISCUSSION

Although surgical competence is a multimodal function, proficiency in technical skills to perform an operative procedure is fundamental to a successful outcome.^{22–24} Assessment within the operating theater is not only a mode of credentialing individual surgeons but also enables collective audit of surgical units and residency training programs.²⁵ Despite the development of a number of tools for assessment of technical skills, none has been incorporated into standard practice. This is either due to their complexity, poor validity, or the lack of experienced personnel to administer them. The only way to ensure data are collected for every single operation performed within a hospital is to develop a system that automatically records and analyzes the required information, without causing any delay or difficulty to the operating room procedure.

TABLE 4. Results From Motion Tracking Parameters, Divided Into Individual Tasks

	Time Taken (s)			Total Path Length (meters)			Total No. Movements		
	Novice (n = 14)	Experienced (n = 33)	<i>P</i>	Novice (n = 14)	Experienced (n = 33)	<i>P</i>	Novice (n = 14)	Experienced (n = 33)	<i>P</i>
Whole procedure	2175 (1954–3127)	1979 (1137–2582)	0.036*	440 (391–565)	423 (274–667)	0.625	1708 (1599–2072)	1771 (1015–2303)	0.389
Dissection of Calot's triangle	854 (768–1056)	393 (243–691)	0.002*	138 (107–196)	73 (39–167)	0.048*	640 (528–866)	367 (197–583)	0.007*
Clip and cut cystic duct	55 (40–105)	33 (18–57)	0.013*	8 (3–15)	5 (2–7)	0.063	23 (14–42)	21 (10–35)	0.553
Clip and cut cystic artery	37 (27–104)	21 (13–30)	0.002*	4 (3–12)	3 (2–5)	0.119	18 (8–33)	13 (8–17)	0.204
Dissection of gallbladder from liver bed	401 (233–837)	374 (207–620)	0.471	75 (61–141)	69 (51–132)	0.377	351 (250–761)	325 (227–541)	0.396

Values are given as medians according to each group, with the interquartile range in parentheses. *P* values are based upon intergroup comparisons from the Mann-Whitney test. *Statistically significant.

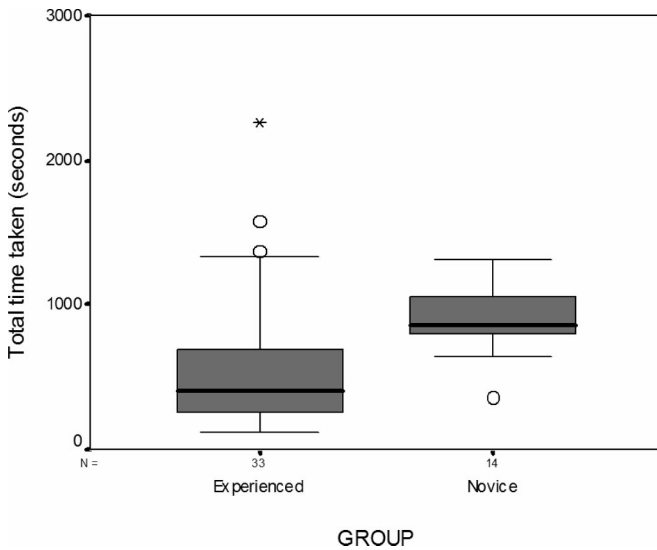


FIGURE 1. Time taken for experienced and novice surgeons to dissect Calot triangle. There was a significant difference between experienced and inexperienced groups ($P = 0.002$).

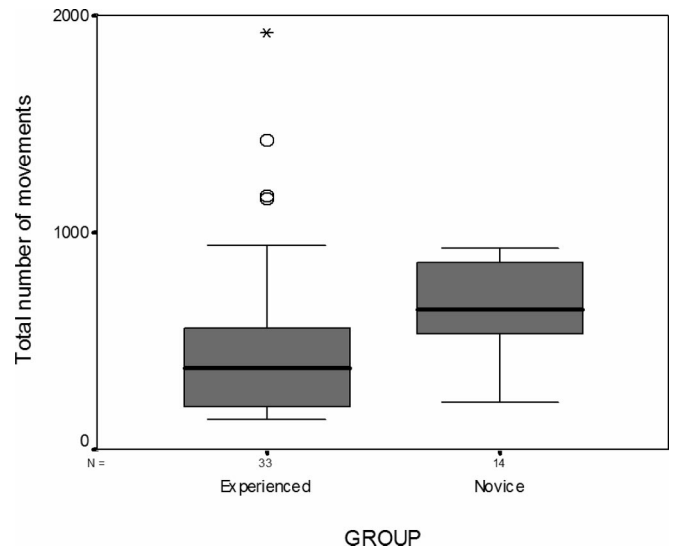


FIGURE 3. Total number of movements for experienced and novice surgeons to dissect Calot triangle. There was a significant difference between experienced and inexperienced groups ($P = 0.007$).

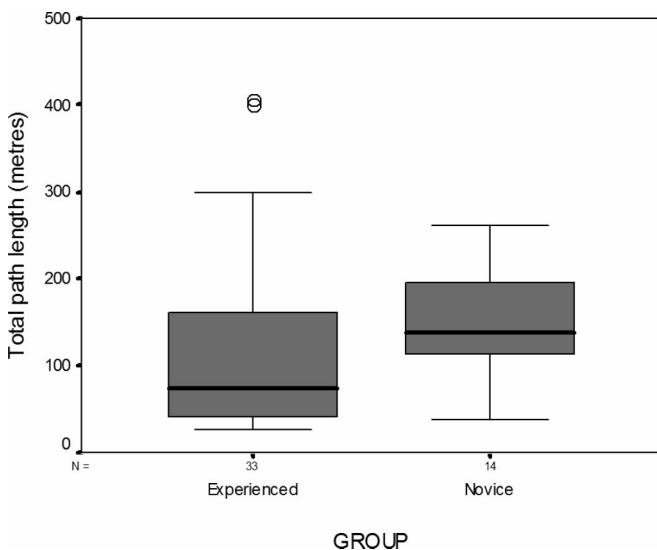


FIGURE 2. Total path length for experienced and novice surgeons to dissect Calot triangle. There was a significant difference between experienced and inexperienced groups ($P = 0.048$).

This was our intention with the development of the ROVIMAS motion tracking system. With this study, we have reiterated the feasibility and confirmed the validity and intertest reliability of this device for assessment of laparoscopic technical skills in the operating theater. In terms of validity, time taken was the only marker to differentiate procedural performance between groups of experienced and novice laparoscopic surgeons. However, the synchronization feature of the system enabled motion tracking parameters to reveal significant differences in dexterity between the 2 groups of surgeons during dissection of Calot triangle. The novice

surgeons on average were twice as slow and half as dexterous as the experienced group. This may be because it is the most difficult part of the operation, and indeed the most likely to lead to a catastrophic error.

Standardization of the procedures was performed on the basis of the patient history, preoperative investigations, and intraoperative findings. A closer inspection of this process reveals that standardization is primarily based upon the degree of inflammation at Calot triangle. It is thus not surprising that this part of the procedure gleaned significant differences in dexterity between the 2 groups of surgeons. None of the other parts of the operation demonstrated construct validity for assessment of dexterity parameters. Reasons for this are either due to simplicity of the task, eg, clip and cut, or anatomic variations such as length of the gallbladder attached to the liver bed. It may also be possible that the failure to achieve significance in dexterity parameters for the whole procedure and other parts of the operation between the 2 groups is due to an underpowered study. Although the intended 20 cases per group were recruited, 6 were excluded upon the basis of intraoperative characteristics.

Nonetheless, it is of concern to note that there remained similar degrees of variability within the novice and experienced groups in terms of dexterity assessments. It would be expected, and indeed has been shown in the literature, that experienced surgeons display a greater degree of consistency when compared with their junior counterparts.²⁶ The conflicting factor may be that surgeons performed the procedure with their “usual technique,” perhaps explaining the variability within the experienced group. It would be necessary to confirm this by dexterity analysis of different techniques to perform laparoscopic cholecystectomy, eg, blunt/sharp versus blunt/teasing dissection methods.²⁷

A significant limitation of dexterity-based assessment using motion analysis is a failure to capture the qualitative

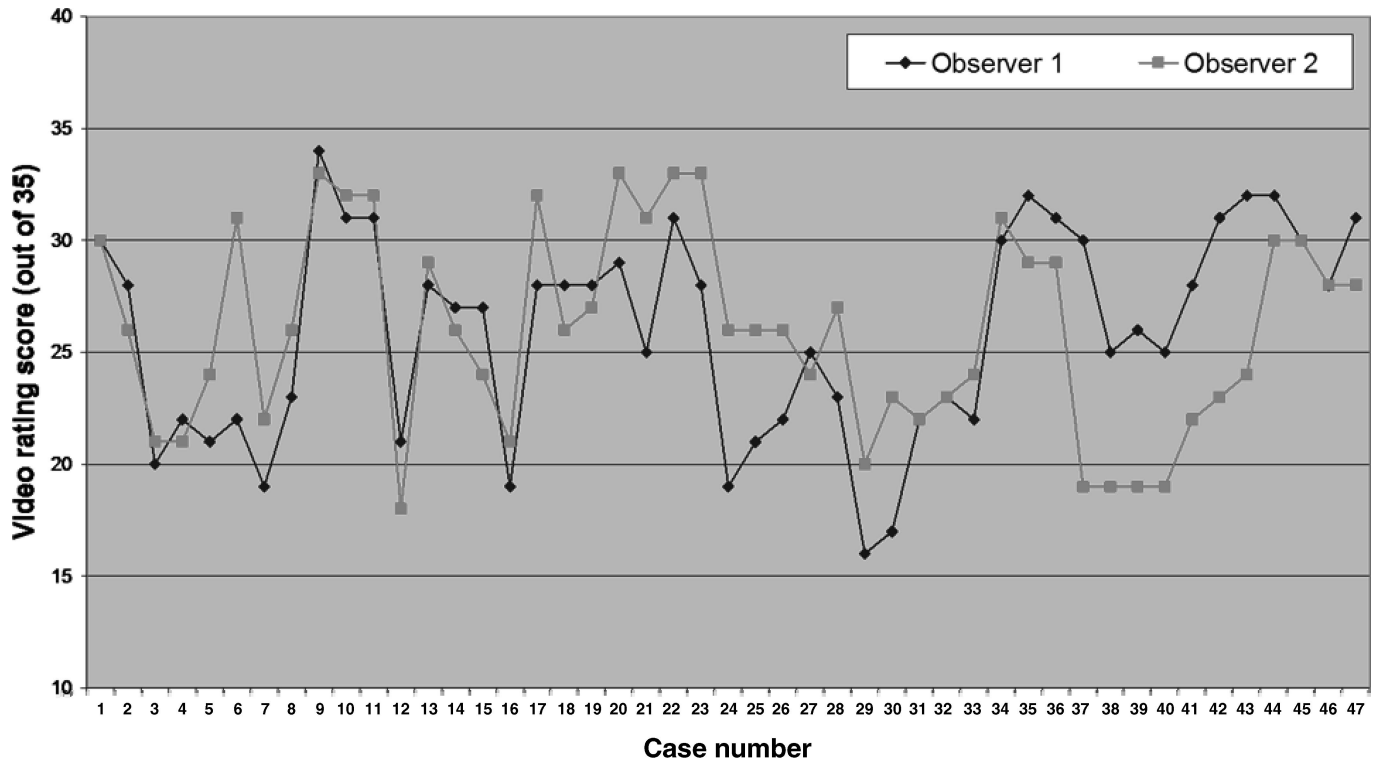


FIGURE 4. Inter-rater reliability of OSATS global rating score between observer 1 and 2 (Cronbach's $\alpha = 0.72$).

TABLE 5. Correlation Between OSATS Global Rating Score and Motion Tracking Data, Calculated Using Spearman Rank Correlation Test

	Correlation With Video Score (r value)	Correlation With Video Score (P)
Whole procedure		
Time taken (s)	-0.625	<0.001
Calot's triangle		
Time taken (s)	-0.468	<0.001
Total path length (ms)	-0.411	<0.005
Total no. movements	-0.414	<0.004
Clip/cut duct		
Time taken (s)	-0.603	<0.001
Clip/cut artery		
Time taken (s)	-0.625	<0.001

and procedural aspects of an operation. Although number of movements can be used as a measure of operative dexterity, performance is more readily measured by number of faulty or inappropriate movements. This was the reason for integration of a video-based analysis of technical skill using the OSATS global rating scale.²⁰ Although other rating scales exist, divided broadly into checklists and error-scoring systems, the OSATS global rating scale has repeatedly been validated for skills assessment. In our study, the generic OSATS scale displayed construct validity. The aim of a global rating scale

is to assess general surgical principles, whereas checklist-based assessments are by definition specific to the operation. Checklists enable detailed evaluations by specifying individual steps and substeps of the operative procedure.²⁸ This is time-consuming in terms of assessment, and awards the surgeon only if they perform the procedure in the predefined sequence of steps. However, surgery is not a mechanical process and thus it is difficult to justify its evaluation in such a rigid manner. The criticism is that such a scale can only ensure whether something was done or not, but not whether it was done well or poorly.

A number of studies have made use of the generic OSATS scale within the operating theater,²⁹⁻³² although the architects of the tool themselves are concerned that "critical aspects of technical skill are not assessed."²¹ Furthermore, this scale was developed for use in live rather than video-based assessment. The Toronto group have subsequently developed and validated a video-based, procedure-specific objective component rating scale for Nissen fundoplication.²¹ In the same vein, researchers have recently sought to develop procedure-specific global rating scales, although their efficacy is yet to be tested.^{33,34}

Joice et al developed an error-based approach to surgical skills assessment that uses human reliability analysis (HRA), which is the systematic assessment of human-machine systems and their potential to be affected by human error.²⁷ Tang et al have applied an Observational Clinical Human Reliability Assessment (OCHRA) tool to laparoscopic cholecystectomy and pyloroplasty.³⁵⁻³⁸ Observational videotape data are subjected to a detailed step-by-step anal-

ysis of surgical operative errors, which are divided into consequential or inconsequential. This is a highly specialized and labor-intensive task, although it has been suggested that OCHRA provides a comprehensive objective assessment of the quality of surgical operative performance by documentation of errors, stage of the operation when they are most frequent, and when they are consequential. In a not dissimilar manner, the use of dexterity analysis has enabled identification of Calot triangle dissection as part of the operation in which there are significant differences between novice and experienced surgeons. However, OCHRA benefits assessment in a more formative manner, enabling error modes to be studied and corrective actions to be pursued. Part of our further work is to define the relative roles of motion analysis, rating scales and the OCHRA tool for surgical skills assessment.

A further question that has rarely been discussed in terms of surgical skills research is the intertest reliability of an instrument for assessment of technical skill.³⁹ The reliability of the assessment was good for motion tracking parameters and excellent for the video-based global rating scale. This adds further weight to the use of these modes of assessment to assess improvements in performance of a trainee surgeon, or the consistency of an experienced surgeon.

It seems that reliable and valid assessment of laparoscopic skills within the operating theater can be performed with ROVIMAS motion tracking software, though not exclusively. The global rating scales were valid and demonstrated higher intertest reliabilities. It is not surprising then that a comparison of motion tracking data with video-based rating scales revealed significant correlations with the global rating scales. An assumption is that the automated motion tracking device can do the work of 2 experienced observers assessing an operation on a global rating scale. Although not true at present, this is certainly a notion we are working toward. With regard to dissection of Calot triangle, significant differences were noted for *both* path length and number of movements. Although both parameters are related, it is of course possible to perform a task with very fine movements, leading to a large number of movements and shorter path length. Current work seeks to define the relationship between these 2 parameters, ie, average path length per movement; a low ratio would suggest fine movements. This information may be useful to highlight accuracy, or perhaps uncertainty, during video-based assessment of the surgical procedure.

However, observer-based assessment of technical skill is time-consuming, and relies upon the availability of experienced surgeons to rate performance.²¹ The dexterity parameters from the motion tracking device may be useful as a first-pass filter, avoiding the need to view the entire procedure to obtain information regarding technical proficiency. In this manner, surgeons could be automatically assessed on motion analysis each time they performed a procedure, and parts of the video rated either by global or error-based scoring systems only if their dexterity parameters fell outside a predetermined range of values. This could reduce the time taken to assess proficiency and would lead to the development of an accurate record of operative skill. Furthermore, many surgeons already make a video record of laparoscopic proce-

dures that they perform; the association of motion tracking data could be stored in a similar operative library.

Extending this concept further leads onto the notion of an operating room black box whereby all aspects of performance are recorded onto a single platform.⁴⁰ As well as a documentation of the surgeon's technical skills, the record would include information regarding the rest of the operative team, the patient, equipment used, etc. As in the airline black box, all of this information could be recorded in an automated manner and available for review at a later stage.

CONCLUSION

The field of surgical skills assessment has progressed significantly over the past decade. To transfer this knowledge from research laboratories and onto real cases, there is not only a need for the further development of these tools but also to define the structure of training programs.^{41–43} Although the concept of competency-based training is regarded as essential to the future of surgical practice, details regarding its implementation have failed to be considered. The primary aim should be to integrate objective assessment of surgical skill into training programs and ensure substandard performance results in remediation or repetition of part of the program. This system can only be warranted with significant financial resources, enabling not only quality assurance of surgical practice, but also additional research into further developing the present tools for comprehensive surgical skills assessment.

ACKNOWLEDGMENTS

The authors thank all surgeons and patients who agreed to participate in the study, and the operating room staff who were involved.

REFERENCES

1. Cuschieri A. Medical error, incidents, accidents and violations. *Min Iniv Ther Allied Technol.* 2003;12:111–120.
2. Reason J. *Human Error.* Cambridge: Cambridge University Press; 1994.
3. Smith R. All changed, changed utterly: British medicine will be transformed by the Bristol case. *BMJ.* 1998;316:1917–1918.
4. Elwyn G, Corrigan JM. The patient safety story. *BMJ.* 2005;331:302–304.
5. Dent TL. Training, credentialing, and granting of clinical privileges for laparoscopic general surgery. *Am J Surg.* 1991;161:399–403.
6. European Association of Endoscopic Surgeons. Training and assessment of competence. *Surg Endosc.* 1994;8:721–722.
7. Jackson B. *Surgical Competence: Challenges of Assessment in Training and Practice,* 5th ed. London: RCS & Smith and Nephew; 1999.
8. Ribble JG, Burkett GL, Escovitz GH. Priorities and practices of continuing medical education program directors. *JAMA.* 1981;245:160–163.
9. Society of American Gastrointestinal Surgeons. Granting of privileges for laparoscopic general surgery. *Am J Surg.* 1991;161:324–325.
10. Moorthy K, Munz Y, Sarker SK, et al. Objective assessment of technical skills in surgery. *BMJ.* 2003;327:1032–1037.
11. Cuschieri A, Francis N, Crosby J, et al. What do master surgeons think of surgical competence and revalidation? *Am J Surg.* 2001;182:110–116.
12. Vincent C, Moorthy K, Sarker SK, et al. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg.* 2004;239:475–482.
13. Reznick RK. Teaching and testing technical skills. *Am J Surg.* 1993;165:358–361.
14. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. *Br J Surg.* 2004;91:1549–1558.

15. Kneebone RL, Nestel D, Moorthy K, et al. Learning the skills of flexible sigmoidoscopy: the wider perspective. *Med Educ.* 2003; 37(suppl 1):50–58.
16. Darzi A, Smith S, Taffinder N. Assessing operative skill: needs to become more objective. *BMJ.* 1999;318:887–888.
17. Dosis A, Aggarwal R, Bello F, et al. Synchronized video and motion analysis for the assessment of procedures in the operating theater. *Arch Surg.* 2005;140:293–299.
18. Hanna GB, Cuschieri A. Influence of two-dimensional and three-dimensional imaging on endoscopic bowel suturing. *World J Surg.* 2000;24: 444–448.
19. Smith SG, Torkington J, Brown TJ, et al. Motion analysis. *Surg Endosc.* 2002;16:640–645.
20. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997;84: 273–278.
21. Dath D, Regehr G, Birch D, et al. Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. *Surg Endosc.* 2004;18:1800–1804.
22. Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. *N Engl J Med.* 1991;324:1073–1078.
23. Hall JC, Ellis C, Hamdorf J. Surgeons and cognitive processes. *Br J Surg.* 2003;90:10–16.
24. Spencer F. Teaching and measuring surgical techniques: the technical evaluation of competence. *Bull Am Coll Surg.* 1978;63:9–12.
25. Urbach DR, Baxter NN. Does it matter what a hospital is ‘high volume’ for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *BMJ.* 2004;328: 737–740.
26. Mackay S, Morgan P, Datta V, et al. Practice distribution in procedural skills training: a randomized controlled trial. *Surg Endosc.* 2002;16:957–961.
27. Joice P, Hanna GB, Cuschieri A. Errors enacted during endoscopic surgery: a human reliability analysis. *Appl Ergon.* 1998;29:409–414.
28. Cao CG, MacKenzie CL, Ibbotson JA, et al. Hierarchical decomposition of laparoscopic procedures. *Stud Health Technol Inform.* 1999;62:83–89.
29. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg.* 2004;91:146–150.
30. Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med.* 1998;73:993–997.
31. Scott DJ, Rege RV, Bergen PC, et al. Measuring operative performance after laparoscopic skills training: edited videotape versus direct observation. *J Laparoendosc Adv Surg Tech A.* 2000;10:183–190.
32. 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;191:272–283.
33. Larson JL, Williams RG, Ketchum J, et al. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery.* 2005;138:640–647.
34. Sarker SK, Chang A, Vincent C, et al. Technical skills errors in laparoscopic cholecystectomy by expert surgeons. *Surg Endosc.* 2005; 19:832–835.
35. Tang B, Hanna GB, Joice P, et al. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Arch Surg.* 2004; 139:1215–1220.
36. Tang B, Hanna GB, Bax NM, et al. Analysis of technical surgical errors during initial experience of laparoscopic pyloromyotomy by a group of Dutch pediatric surgeons. *Surg Endosc.* 2004;18:1716–1720.
37. Tang B, Hanna GB, Cuschieri A. Analysis of errors enacted by surgical trainees during skills training courses. *Surgery.* 2005;138:14–20.
38. Tang B, Hanna GB, Carter F, et al. Competence assessment of laparoscopic operative and cognitive skills: Objective Structured Clinical Examination (OSCE) or Observational Clinical Human Reliability Assessment (OCHRA). *World J Surg.* 2006;30:527–534.
39. Bann S, Davis IM, Moorthy K, et al. The reliability of multiple objective measures of surgery and the role of human performance. *Am J Surg.* 2005;189:747–752.
40. Guerlain S, Adams RB, Turrentine FB, et al. Assessing team performance in the operating room: development and use of a ‘black-box’ recorder and other tools for the intraoperative environment. *J Am Coll Surg.* 2005;200:29–37.
41. Darosa DA. It takes a faculty. *Surgery.* 2002;131:205–209.
42. Debas HT, Bass BL, Brennan MF, et al. American Surgical Association Blue Ribbon Committee Report on Surgical Education: 2004. *Ann Surg.* 2005;241:1–8.
43. Pellegrini CA, Warshaw AL, Debas HT. Residency training in surgery in the 21st century: a new paradigm. *Surgery.* 2004;136:953–965.