

A Human Factors Analysis of Technical and Team Skills Among Surgical Trainees During Procedural Simulations in a Simulated Operating Theatre

Krishna Moorthy, MS, MD, FRCS, Yaron Munz, MD, Sally Adams, PhD,
Vikas Pandey, BSc, MBBS, FRCS, and Ara Darzi, MD, FRCS, FACS

Background: High-risk organizations such as aviation rely on simulations for the training and assessment of technical and team performance. The aim of this study was to develop a simulated environment for surgical trainees using similar principles.

Methods: A total of 27 surgical trainees carried out a simulated procedure in a Simulated Operating Theatre with a standardized OR team. Observation of OR events was carried out by an unobtrusive data collection system: clinical data recorder. Assessment of performance consisted of blinded rating of technical skills, a checklist of technical events, an assessment of communication, and a global rating of team skills by a human factors expert and trained surgical research fellows. The participants underwent a debriefing session, and the face validity of the simulated environment was evaluated.

Results: While technical skills rating discriminated between surgeons according to experience ($P = 0.002$), there were no differences in terms of the checklist and team skills ($P = 0.70$). While all trainees were observed to gown/glove and handle sharps correctly, low scores were observed for some key features of communication with other team members. Low scores were obtained by the entire cohort for vigilance. Interobserver reliability was 0.90 and 0.89 for technical and team skills ratings.

Conclusions: The simulated operating theatre could serve as an environment for the development of surgical competence among surgical trainees. Objective, structured, and multimodal assessment of performance during simulated procedures could serve as a basis for focused feedback during training of technical and team skills.

(*Ann Surg* 2005;242: 631–639)

Surgical training has responded to the challenges of reduced training opportunities, shortened working hours, and financial pressures by using simulations primarily for the acquisition of psychomotor skills. Simulations are used to augment training in the operating room (OR) and trainees acquire their skills in a “nonthreatening and unhurried environment.”¹ Learning in the operating theater provides little opportunity for practice and reflection, and this is one of the strongest stimuli for taking training in surgical skills into non-OR environments.^{2,3} Task simulations possess the potential to structure the learning of technical skills in a manner aimed at ensuring a smooth escalation in task complexity leading to the performance of procedures in the operating theater.⁴ However, even such a structured transition from the skills laboratory to the operating theater fails to ensure that technical skills learning is linked to the training of other skills such as gloving and gowning, adherence to asepsis, and interpersonal skills such as communication and leadership. While simulation based learning of technical skills may allow for learning and practice in a controlled environment, an approach focusing solely on technical skills can ignore these skills that underpin surgical competence.

In addition, the assessment of technical proficiency in surgery has been restricted to the observation of task performance on bench and animal models in a laboratory setting.^{5,6} Although such methods go a long way in making the assessment of technical skills more objective, they still suffer from the drawback that these assessments are conducted in non-OR environments.

Training in all specialties of medicine in addition to surgery places a large emphasis on technical and clinical skills alone. This is similar to the position the aviation industry was in about 20 years ago. The focus was on technological developments to the aircraft and on technical training of pilots until research sponsored by the National Aeronautics and Space Administration revealed that 70% of errors were due to human causes. NASA’s research revealed that a large number of errors were a result of failures in interpersonal communication, decision making, and leadership.⁷ The similarities that exist between aviation and the medical specialties have already prompted certain specialties such as anesthesia and intensive care to adopt some of aviation’s training strategies.⁸

From the Imperial College- St. Mary’s Hospital Simulation Group, Department of Surgical Oncology and Technology, Imperial College, London. The Imperial College-St. Mary’s Hospital Simulation Group includes the following: Shirley Martin, Lorraine Poore, Lee Edwards, John Abbot, Janet Henry, Benny Lo, Prof. Gunag Z. Yang, and Prof. Charles Vincent. Supported in part by a grant obtained from the BUPA Foundation.

Reprints: Krishna Moorthy, MD, 28, Carless Avenue, Birmingham, B17 9EQ, UK. E-mail: k.moorthy@imperial.ac.uk.

Copyright © 2005 by Lippincott Williams & Wilkins
ISSN: 0003-4932/05/24205-0631

DOI: 10.1097/01.sla.0000186298.79308.a8

While surgical trainees receive feedback on their technical performance from their trainers, however, subjective and non-criteria based it may be,¹ surgical trainees rarely ever receive any feedback on their nontechnical or team skills. Communication skills training consist solely of training in doctor-patient skills and not interprofessional skills. The importance of nontechnical skills is highlighted by the fact that errors in the operating theater are rarely due to deficiencies in technical performance.⁹ They result more from impaired decision-making, absence of situation awareness, and failures in interpersonal communication. Any training strategy that aims to enhance team working in the OR should improve performance in the OR and thus reduce the incidence of errors.

The aims of this study were to develop a simulated operating theater environment and establish its feasibility and face validity, establish the construct validity and reliability of the methods used for the assessment of technical skills, develop and explore the reliability of a nontechnical skills assessment method and pilot such a method to observe nontechnical skills in surgeons, and also explore any differences between grades of trainees.

MATERIALS AND METHODS

Simulated Operating Theatre (SOT)

Physical Environment

A simulated operating theater (SOT) (Fig. 1a) was developed with the aim of replicating a real operating theater as closely as possible with an adjacent room serving as a control room (Fig. 1b). A novel technology called the Clinical Data Recorder (CDR) was developed to evaluate the technical skills of the surgeon as well as the communication and interaction between OR personnel. The CDR allows multiple streams of video and audio data to be collected and recorded onto a Digital Versatile Disc (DVD) recording device in an encrypted format.

The researchers and trainers view the proceedings in the SOT on a monitor placed in the control room. The data from the cameras is initially fed into a video-mixer before being viewed on the monitor. The images on the monitor thus depend on the number and the order of the cameras selected. For the purposes of this project, we used only 4 cameras, the input from which was fed into the CDR. The audio from the SOT is streamed into the monitor and the recording medium. Thus, the researcher/trainers are able to view the simulation in real time as well as record the simulation onto a DVD disc for future review and evaluation.

Equipment

In addition to all the standard operating theater equipment, one of the primary features of the SOT is an anesthetic simulator (SimMan, Laerdal, UK) (Fig. 1a). This is a simulator of moderate fidelity, which allows manipulation of the mannequin's hemodynamic parameters through software installed on a notebook computer located in the control room. It is possible to create a number of scenarios using the software. These include hypoxia, laryngospasm, and pneumothorax, etc. The mannequin can also be intubated and attached to the anesthetic machine.



FIGURE 1. Simulated Operating Theatre (SOT) with the anesthetic simulator (SimMan, Laerdal, UK). The SOT is separated from the control room, shown in the lower figure, by a one-way glass. The trainers/raters watch the simulation on monitors placed in the control room.

Subjects

Surgical trainees were divided into 3 groups according to their level of experience.

Group 1: Junior trainees (less than 20 saphenofemoral junction high tie procedures).

Group 2: Middle level trainees (20–50 procedures).

Group 3: Senior trainees (>50 procedures).

The groups reflect the average number of procedures performed by 3 groups of surgical trainees in the UK, namely, basic surgical trainees, junior higher surgical trainees, and senior higher surgical trainees.

Procedure

A previously validated synthetic model of the saphenofemoral junction (Limbs and Things, Bristol, UK) was fixed to the anesthetic simulator. This is a silicon-based model with accurate simulation of the saphenofemoral junction with a layer of simulated skin overlying a layer of superficial fascia. It consists of a “saphenous vein” with 4 tributaries connected to a “femoral” vein. These are set in a cast of silicon, which resembles fat. The model was then draped with surgical drapes and held with surgical towel clips.

TABLE 1. Imperial College Evaluation of Procedure-Specific Skills (ICEPS)

	1	2	3	4	5
Incision	Does not use surface landmarks Inappropriate placement of incision Poor handling of scalpel		Appropriate incision in terms of location and size Looked at ease with scalpel		Handled scalpel expertly
Dissection	Appeared unsure and excessively hesitant while dissecting Caused trauma to tissues Did not dissect into the correct anatomic plane		Controlled and safe dissection into correct anatomic plane Caused minimal trauma of tissues Used instrument satisfactorily while dissecting		Superior and atraumatic dissection into the correct anatomic plane Confident handling of instruments while dissecting
Retraction	Clumsy use of retractors Did not allow visualization of important structures making frequent changes to retractor setting		Good use of retraction allowing visualization of major structures Had to change retractor position to visualize other structures		Excellent use of retractors Allowed good visualization of all necessary structures Atraumatic
Tributaries	Could not or did not try to identify any tributaries		Identified all known tributaries Did not seek other vessels		Identified all known tributaries Sought other possible tributaries
Hemostasis	Poor quality of knot tying Knots frequently slipped or were excessively traumatic to vessels		Competent knot tying Minimal trauma to vessels Minimal blood loss		Superior knot tying Atraumatic No knot slippage
SFJ clearance	Did not identify the SFJ or excessively traumatic dissection around that vessel		Identified the SFJ Safely dissected tissues away from vessel Reasonable clearance of vessel Minimal trauma		Identified the SFJ Expert dissection of tissues off the vessels Atraumatic Cleared well proximally and distally
SFJ ligation	Did not ligate the SFJ or ligated CFV or caused excessive encroachment onto CFV after SFJ ligation		Good knot tying while ligating the SFJ Minimal encroachment onto CFV following SFJ ligation		Excellent safe and secure ligation of the SFJ Flush ligation with no encroachment onto CFV

SFJ indicates saphenofemoral junction; CFV, common femoral vein.

Simulation

A standardized theater team consisting of an anesthetist, an operating department assistant, a scrub nurse, a “circulating” nurse (assistant to the scrub nurse), and an assistant first entered the SOT and assumed their positions. The surgeons entered the simulation area through the control room adjacent to the operating theater. They were briefed about the scenario and made familiar with the theater environment through the one-way glass. They were then asked to sign a consent form for participating in the study with a clause requesting their confidentiality. They then entered the operating theater under the assumption that they had scrubbed and proceeded to gown and glove.

While performing the simulations, there was a preprogrammed hypoxia scenario that was created to assess the awareness of the surgeon toward the patient’s condition and the anesthetist. This was done by the simulation controller/research fellow by manipulating the software of the anesthetic simulator. The anesthetists were requested not to prompt the surgical trainee about the crisis, but were asked to do so when prompted by the simulation controller. This was usually communicated as a signal by the simulation controller when the surgical trainee had failed to acknowledge the existence of a problem. All the simulations were recorded onto a DVD disc for subsequent analysis by independent observers.

Post-Simulation Feedback/Debriefing

This was given to the participants subsequent to the simulation. They were given a feedback of their technical skills by a surgeon and nontechnical skills by a human factors expert and a trained research fellow. These were all video-based sessions. The simulations were time-marked by the assessors during the assessment process, and these segments were played back during the feedback session.

Performance Measures and Validity of the Simulation

Assessment of Technical Skills

A close-up image of the procedure was shown to 3 independent blinded observers instructed in the use of a global rating scale developed by Reznick et al: Objective Structured Assessment of Technical Skills (OSATS).¹⁰ This consists of different aspects of surgical skills rated on a 5-point scale with the middle and the extremes anchored by explicit descriptors.⁵

In addition to the generic global rating scale, they also assessed technical performance using a procedure-specific global rating scale: Imperial College Evaluation of Procedure-Specific Skills (ICEPS).¹¹ This consists of 7 steps of the procedure, saphenofemoral junction high-tie, which are rated in a manner similar to the global rating scale (Table 2).

TABLE 2. Essential Items Checklist

Step	Yes	No
Gowning		
Gloving		
Checking consent and side with team		
Patient positioning (Trendelenburg)		
Asking anesthetist to start		
Handling of sharps (safely)		
Adhering to asepsis throughout the procedure		
Waiting for swab/instrument check prior to closure		
Acknowledging swab count		
Informing anesthetist of closing		

Essential Item Checklist

This consisted of parameters (Table 1) that are critical to safe and effective performance in the operating theater. Some of these parameters are embedded in the technical skills assessment (Op Comp) guide for trainers as published by the Surgical Advisory Committee in general surgery in the United Kingdom.¹² Each item in the checklist was analyzed separately. One independent observer (a surgeon) carried out the checklist assessment. This checklist consisted of 4 essential and observable aspects of team communication, which are considered to be important during an operative procedure.¹³

Assessment of Nontechnical Skills

A few elements of the LOSA checklist,⁷ developed for the assessment of nontechnical skills in aviation which were considered to be relevant to surgery, were used with the intention of exploring the feasibility and interrater reliability of nontechnical skills assessment. These consisted of the rating of behaviors such as preoperative preparation, communication, vigilance (situation awareness), and leadership. These were then subdivided into elements (Table 3) that were rated on a 5-point Likert scale with the 2 extremes being anchored by descriptors. The total score was then expressed as a percentage. These measures were also based on the results of a pilot questionnaire submitted to a group of 35 consultants (attending physicians) and higher surgical trainees. The nontechnical skills assessment was carried out by 2 independent assessors: a human-factors expert with experience in simulations in the nuclear industry and the research fellow who had been trained by the former. During the training period, the first 5 assessments were conducted together to ensure uniformity of assessment. Subsequently, the assessments were performed independently to permit the evaluation of interrater reliability.

Communication: Utterance Frequency (UF)

In addition a communication count (UF) was performed to assess the communication of the surgeon with other team members. This method is similar to an approach used in an earlier study observing communication in the operating theater.¹⁴ The UF was defined as the number of episodes of communication per minute. A total UF was then calculated for the communication between the surgeon (S) and the

TABLE 3. Nontechnical Skills Assessment

Behavior	Element
Preoperative preparation	(a) Introduction to team members
	(b) Preoperative instrument and equipment check
	(c) Briefing
Communication and interaction	(a) Instructions to assistant/scrub nurse: clear and polite
	(b) Awaits acknowledgment from the assistant/scrub nurse
	(c) Assistance sought from team members
	(d) Acknowledges help/advice from team members
Vigilance/situation awareness	(a) Monitored patient's parameters throughout the procedure
	(b) Awareness of anesthetist
	(c) Actively initiates communication with anesthetist
Leadership	(a) Adherence to best practice during the procedure
	(b) Resource utilization, ie, appropriate task-load distribution and delegation of responsibilities
	(c) Authority/assertiveness

anesthetist (An), assistant (A) and the scrub nurse (SN). This is described as S-An, S-A, and S-SN, respectively.

Participant Assessment of the Realism of the Simulation: Face Validity

The participants were asked to respond to 4 statements as disagree, not sure and agree to establish the face validity and the training value of the simulation.

The SOT was as realistic as a real operating theatre.

The model was a realistic representation of a real procedure.

The simulated procedure in the SOT is a good method for training technical skills.

The simulated procedure in the SOT is a good method of training nontechnical skills.

Data Analysis

As the data were not found to have normal distribution for most measures, nonparametric tests were used for analysis. Differences in between groups for the technical and nontechnical skills were done using the Kruskal-Wallis test. The interrater reliability (the level of agreement between the observers) was determined by using Cronbach's alpha coefficient. χ^2 test was used for analysis of the essential items checklist data. Spearman's Rank correlation test (coefficient rho) was also used to determine the correlation between the technical skills score and the nontechnical score and between the communication score of the nontechnical skills assessment and the UF count.

RESULTS

There were 27 surgeons who participated in the study. Informed consent was taken from all the partici-

pants. There were 11 trainees in group 1 (junior), 9 in group 2 (intermediate), and 7 in group 3 (senior). There were 21 males and 6 females and all the participants were right-handed.

Technical Skills Assessment

Interrater Reliability for the Technical Assessment

The interobserver reliability (alpha coefficient) between the 3 expert observers for the global rating scale was 0.90, and for the ICEPS score was 0.94.

Difference Between the Groups: Construct Validity

There was a significant difference in the global score across the whole range of trainees ($P = 0.002$) (Fig. 2). While the difference in the global score between groups 1 and 2 was significant ($P = 0.01$), the difference between groups 2 and 3 was not significant ($P = 0.31$), although group 3 trainees did score higher.

There was a significant difference across the 3 groups for the ICEPS score as well ($P = 0.002$) (Fig. 2). While there was a significant difference in the scores between groups 1 and 2 ($P = 0.012$), this was not observed for the scores between 2 and 3 ($P = 0.56$), although group 3 did score higher than group 2.

Essential Items Checklist

Intergroup Differences

All the subjects gowned and gloved correctly and handled sharps safely. There were no differences between the groups for patient positioning ($P = 0.46$), informing anesthetist of start of procedure ($P = 0.43$), asepsis ($P = 0.19$), waiting for the swab count (0.56), acknowledgment of the swab count ($P = 0.36$), and informing the anesthetist upon completion of the procedure ($P = 0.50$).

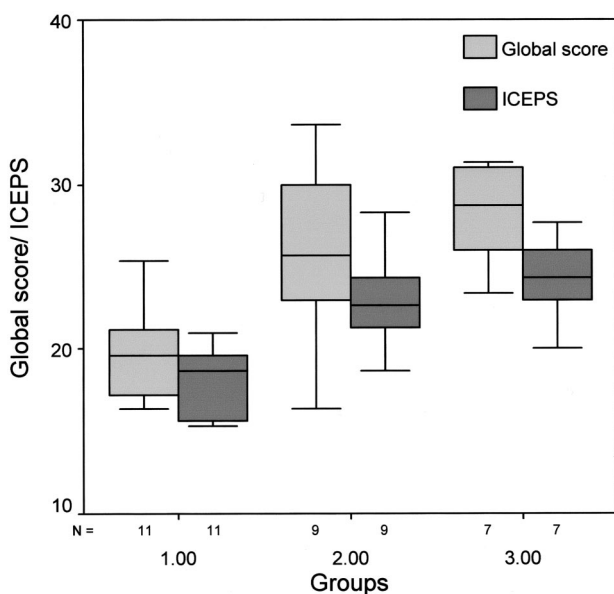


FIGURE 2. Generic global score and ICEPS score for the 3 groups.

Data for the Entire Cohort of Trainees

Analysis of the key communication features revealed that 74% of the entire cohort informed the anesthetist prior to taking the incision, but only 21% informed the anesthetist before proceeding to close the wound. Only 8% of them waited for a swab count from the scrub nurse before starting, but 79% of them acknowledged the swab count when informed; 8% of them positioned the patient in the Trendelenburg position, and 88% of them followed all aseptic precautions. Thus, the poorest scores were observed for patient positioning, waiting for swab count, and informing anesthetist prior to closure.

Nontechnical Skills Assessment

Interrater Reliability for the Nontechnical Assessment

The interrater reliability between the 2 independent observers was 0.84 (alpha).

Intergroup Differences for the Total Nontechnical Skills Score

There was no difference in the total nontechnical global score across the 3 groups of trainees ($P = 0.70$). Figure 3 shows the variability within the groups.

Intergroup Differences for Individual Components

There were no significant differences across the groups for any of the components except for leadership ($P = 0.008$). While there was a significant difference between groups 1 and 2 ($P = 0.003$), the difference between groups 2 and 3 was not significant ($P = 0.54$).

Data for the Whole Cohort

For the whole cohort of subjects, the lowest scores were obtained for preoperative preparation and vigilance. The scores, expressed as percentages, were 35.8% and 45.9%, respectively. The score for leadership for the whole cohort was 56.6%.

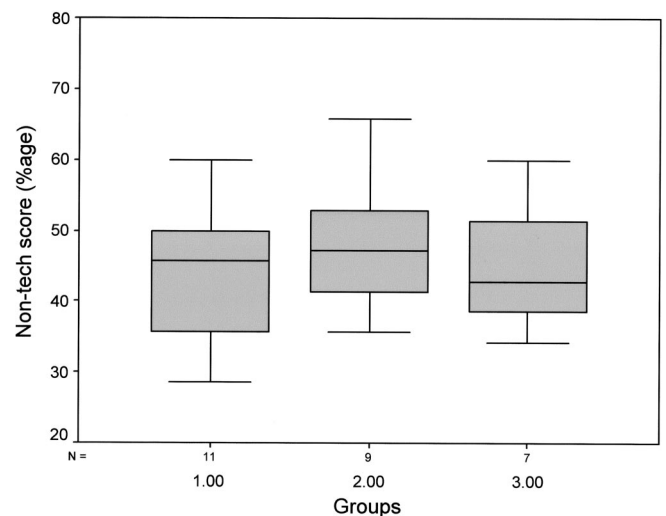


FIGURE 3. Nontechnical skills: percentage score.

Communication: UF

Difference in the Total UF Across the Groups

There was no difference in the total UF across the 3 groups ($P = 0.20$). Table 4 describes the total UF between the surgeons and the individual team members for the 3 groups separately. Analysis between the groups revealed that for the there was a difference in the total UF between the junior trainees and the group with intermediate skill, but this difference was not significant ($P = 0.06$). Similarly, the difference between the senior and the intermediate group was also not significant ($P = 0.47$).

Difference in the UF Between the Surgeon and Individual Team Members

In addition, there was no difference between the 3 groups for the total UF between the surgeon and the assistant (S-A) ($P = 0.47$), between the surgeon and the anesthetist (S-An) ($P = 0.26$) and between the surgeon and the scrub nurse (S-SN) ($P = 0.18$) (Table 4).

There were no differences between groups 1 and 2 for S-A ($P = 0.34$), S-An ($P = 0.19$), and S-SN ($P = 0.06$) UF. There were also no differences between groups 2 and 3 for S-A ($P = 0.56$), S-An ($P = 0.47$), and S-SN ($P = 0.62$) UF.

Data for the Entire Cohort

There was a significant difference in the UF between the surgeons and the individual team members with the UF being significantly lower for S-An compared with S-A and S-SN ($P < 0.001$) (Fig. 4). While there were significant differences between S-A and S-An ($P < 0.001$) and S-SN and S-An ($P < 0.001$), the difference between S-A and S-SN was not significant ($P = 0.09$). This means that the lowest communication occurred between the surgeons and the anesthetists.

Surgeon-Anesthetist UF

A large part of the total communication between the surgeon and the anesthetist was during the hypoxia scenario [prehypoxia UF (median \pm interquartile ratio), 0.24 (0.37); during hypoxia, 2 (2.3); posthypoxia, 0.15 (0.69)].

Correlation Between the UF and the Communication Score in the Nontechnical Assessment

There was a low correlation between the communication score embedded in the nontechnical assessment and the total UF ($\rho = 0.29$) meaning that the 2 measures were assessing different aspects of communication.

TABLE 4. Utterance Frequency Data for the Three Groups

	Total UF	Surgeon- Assistant	Surgeon- Anesthetist	Surgeon- Scrub Nurse
Group 1	3.2 (1.8)	1.5 (0.9)	0.34 (0.36)	1.6 (0.63)
Group 2	4 (2.25)	1.27 (0.6)	0.41 (0.49)	1.9 (1.1)
Group 3	3.3 (2.6)	1.3 (0.9)	0.46 (0.61)	1.5 (1.5)

Median values are shown; values in parentheses are the interquartile range.

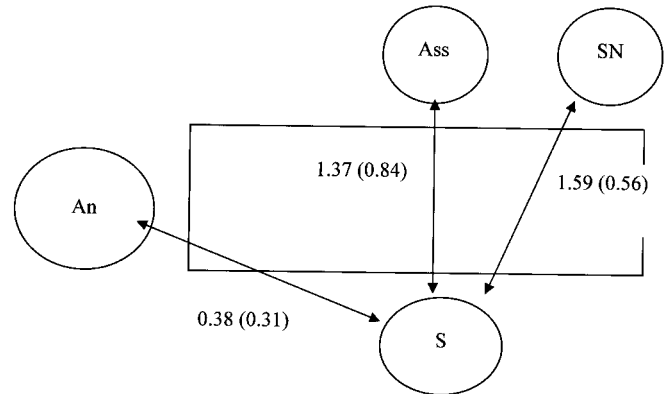


FIGURE 4. Average utterance frequency (UF) of the 3 groups of surgeons with other team members. S, surgeon; An, anesthetist; A, assistant; SN, scrub nurse.

Correlation Between the Technical and Nontechnical Skills

There was a low correlation between the technical and nontechnical global scores of the subjects ($\rho = 0.24$, $P = 0.23$). There was a low correlation between the 2 measures for groups 1 and 2, but there was a moderate correlation for group 3: group 1, $\rho = -0.13$, $P = 0.70$; Group 2, $\rho = 0.29$, $P = 0.44$; Group 3, $\rho = 0.44$, $P = 0.32$.

Face Validity of the Simulation

A total of 90% of the subjects felt immersed in the simulation (suspended disbelief). They agreed with the statement that the SOT was a realistic representation of an operating theatre.

Only 50% of the subjects considered the synthetic model to be a realistic representation of a real procedure.

Debriefing (Feedback)

An effort was made to debrief all the participants, but this was possible for only 17 of the 27 (63%) subjects. This was because of some subjects having moved on to posts at other hospitals. There were 6 junior trainees, 8 of intermediate grade, and 3 of senior grade. All sessions were within 8 weeks of simulations.

A total of 80% of the trainees considered the model to be a good method of training technical skills. The junior trainees (100%) agreed more with this statement compared with the intermediate (75%) and the senior trainees (65%).

A total of 88% of the trainees considered the SOT to be a good environment for training of team/nontechnical skills. Three trainees were not sure.

DISCUSSION

Crew Resource Management training in aviation⁷ and Anesthesia Crisis Resource Management in anesthesia,⁸ attempt to create near-real environments to train and assess technical performance with reference to team or nontechnical skills. In high-risk environments such as the ones mentioned above, individuals work in teams where the performance of one team member can influence the outcome of the entire

mission/procedure. In such circumstances, it is difficult to dissociate technical proficiency from team skills. The primary aim of simulation-based training programs is to integrate technical with team skills training to make the most effective use of all resources available during management of various scenarios.

This study was developed with the intent of adopting a more holistic approach to the learning and assessment of skills. Using a live animal based “wet-laboratory,” Lossing et al described a similar approach that consisted of teaching trainees patient preparation, draping, aseptic techniques, principles of hemostasis, surgical assisting, and overall conduct in the OR.¹⁵ However, the use of live animals was seen as a major drawback of this approach.

Over the past few years, there have been a number of alternatives to operating on live animals such as synthetic bench models and VR simulators. Synthetic bench models have been shown to be effective in the training of technical skills,¹⁶ and using the global rating scale embedded in OSATS, one study has shown a strong correlation between procedures on bench models and live operations using a bench model of a saphenofemoral junction.¹⁷

The global rating scale embedded within OSATS is a generic measure of operative expertise; and even though one of its features assesses knowledge of the procedure, it fails to give detailed procedural information. Procedure-specific global rating scales¹¹ assess knowledge of the procedure on a 5-point Likert scale and combine the advantages of global rating scales with the procedure specific information obtained from checklists. Using both these methods, it was demonstrated that there was a significant difference across the grades but no difference between middle and senior level trainees. Other groups in addition to us have commented on the existence of a “ceiling effect” that is observed during the assessment of technical skills.^{18,19}

One of the reasons for the apparent failure to make nontechnical skills a focus of training among medical specialties has been the difficulty encountered in the objective assessment of these skills to enable structured feedback. This study piloted the use of a simple nontechnical skills assessment scale, which consisted of measures considered important for team performance in the OR from the results of a questionnaire survey undertaken by a small cohort of surgeons. Although this process establishes the content validity of the measures to a limited extent, future research should adopt a cognitive task analysis approach as used for the development of a nontechnical skills assessment for anesthesiologists.²⁰ The nontechnical skills assessment used in this study was found to have a high interobserver reliability. This is probably because there were only a few measures on the scale, resulting in a lower likelihood of variability between the 2 assessors. The benefit of having only a few measures in a nontechnical skills assessment has been suggested by Gaba et al who reasoned that one of the reasons for the low interrater reliability in their observation study of anesthesia simulations was the presence of multiple measures.²¹

The nontechnical skills assessment revealed that there were no differences between the 3 grades for all measures except

for leadership skills. The absence of construct validity could be a result of a number of factors. It could be argued that such skills develop with experience and that a cohort of senior surgeons would have consequently achieved a higher score. Failure to demonstrate this may be due to the fact that the study's cohort consisted only of surgical trainees. Another reason could be that the measures used had little relevance to surgery. However, as stated earlier, we chose these measures according to the results of a pilot questionnaire but acknowledge that further research is required to establish the true content validity of the measures. Most importantly, establishing construct validity of nontechnical skills assessment may be difficult as they have never been the focus of surgical training while a strong emphasis has been placed on technical skills. The low correlation between the technical and nontechnical scores confirms this, although it can be seen that the correlation was highest for the more senior trainees. Variability in performance even within the groups demonstrates that some senior surgical trainees score lower than their junior counterparts as performance depends on a number of complex factors such as mentoring, culture, personality, and exposure to positive role models. It is also important to appreciate that there is very little evidence on the construct validity of nontechnical skills even in aviation and anesthesia.^{21,22}

Data for nontechnical skills for the entire cohort revealed that most trainees scored very low for vigilance. It has been found that a large number of adverse events during the maintenance of anesthesia are due to problems with the vigilance of the anesthetist⁹ and situation awareness training is considered to be an important aspect of anesthesia simulations.²³ This particular condition, however, has never before been studied among surgeons. The lack of awareness of the patient's condition could be either because most surgeons focus solely on completing their task²⁴ or may be a reflection of a cultural norm where they rely on information on the patient's condition from the anesthetist. This needs to be explored by further studies, which should examine the importance of the surgeon's awareness on patient safety in the OR. Error management in aviation places a large emphasis on the fact that team members maintain open lines of communication and cross-check each other.²⁵ One observational study of team coordination emphasized the role of “activity monitoring” where one team member watches another complete their tasks. However, even just observing was found to be problematic, and the authors recommended the need for explicit verbal communication.²⁶

In addition to the assessment of communication as a measure in the nontechnical assessment, UF was also used. The utterance frequency, based on a similar method used by a previous study,¹⁴ was used because of its objectivity. The low correlation between the UF score and the communication score embedded within the nontechnical skills assessment reflects that they were assessing different aspects of communication, with the former being a quantitative measure and the latter assessing the quality of the communication.

It was found that there were no differences between the different grades for the total UF and for the UF between surgeons and individual team members. However, general observational data for the entire cohort revealed some inter-

esting findings. It was observed that communication between the surgeons and anesthesiologists was significantly lower than that with other team members. This is not a surprising finding considering the fact that a majority of the communication between the surgeon and the assistant and the scrub nurse consisted of instructions or requests for instruments. It was also observed that a large component of the communication between the surgeons and the anesthesiologists was during the hypoxia scenario. Of particular interest was the finding that, with a third of the study cohort, there was no communication between the surgeon and the anesthesiologist in the postscenario period, suggesting the lack of vigilance of the patient's condition even after the crisis.

Communication was also assessed for some key elements of communication between team members that is considered to be essential. Although there were no differences between the 3 groups, observations revealed that while the subjects were consistent in asking the anesthesiologist about the appropriateness to start, only one fourth of them informed the anesthesiologist about the completion of the procedure. This is considered to be an essential part of the communication between the surgeon and the anesthesiologist, as it prepares the anesthesiologist to start instituting measures to reverse the state of anesthesia at the optimum time. It was also observed that surgeons proceeded to closing the surgical wound without the swab and instrument count being delivered to them. In both these instances, it may be assumed that activity monitoring between the team members would be adequate, but Xiao et al have advocated the need for training in communication to make "certain verbalizations mandatory" for smooth team coordination.²⁷

The scores for preoperative preparation were low for all 3 groups. This is a reflection of the absence of preoperative briefing, absence of preoperative instrument check, and the inadequacy of introduction to the other team members. The absence of briefings in operating theaters has been noted by earlier studies.²⁸ Using a 4-point scale and comparing communication in the cockpit with that in the OR, Sexton and Helmreich found that, while below standard briefings occurred in only 23% of cockpit observations, they occurred in nearly 90% of OR observations.²⁹ This was despite the fact that in a questionnaire to OR personnel (ORMAQ) a majority of respondents considered preoperative discussion between team members to be important to enhance team working in the OR.³⁰

The face validity of the simulated environment was found to be quite high as nearly 90% of the trainees felt immersed in the simulation. However, the face validity of the model was satisfactory to only 50% of the cohort. In spite of this, 80% of the trainees considered the model suitable for technical skills training. Nearly 90% of the cohort considered the SOT to be suitable environment for team skills training.

One of the limitations of such an approach toward OR training is the lack of suitable models to simulate surgical procedures. While anesthetic simulators, with varying levels of fidelity, are used to train anesthetic personnel, the lack of analogous surgical models places a limitation on the number of procedures that can be simulated. While the development

of virtual reality simulators may address this issue in the future, presently, the SOT training approach could be extended to carrying out procedures on live animals. As live animal-based training is already prevalent in a number of centers in the United States and Europe, it would only involve the shift of technical skills training from skills laboratories to simulated environments such as the SOT. This would be particularly useful for senior trainees (residents) who have already acquired their core technical skills.

In addition to the training and assessment of individual surgical trainees, an SOT-based training approach can be extended to training OR teams. With anesthesiologists already using SOTs for training their personnel, further efforts could be made for surgeons and anesthesiologists to carry out simulations together to enhance team coordination in the OR. Such simulations could also involve other personnel such as nurses and operating department assistants/technicians. Research in our center is presently focused on developing and validating team performance measures¹³ and using skill measures as described in this study to observe performance during real procedures and to determine the impact of individual and team performance on patient outcomes.

ACKNOWLEDGMENTS

The authors thank all the surgical trainees who participated in the study and Mr. S. Bann, Mr. V. Datta, Dr. S. Mackay, and Dr. R. Kneebone with help in the development of the Simulated Operating Theatre.

REFERENCES

1. Reznick RK. Teaching and testing technical skills. *Am J Surg.* 1993; 165:358–361.
2. Hamdorf JM, Hall JC. Acquiring surgical skills. *Br J Surg.* 2000;87: 28–37.
3. Macintyre IM, Munro A. Simulation in surgical training. *BMJ.* 1990; 300:1088–1089.
4. Kaufman HH, Wiegand RL, Tunick RH. Teaching surgeons to operate: principles of psychomotor skills training. *Acta Neurochir (Wien).* 1987;87:1–7.
5. Reznick R, Regehr G, MacRae H, et al. Testing technical skill via an innovative 'bench station' examination. *Am J Surg.* 1997;173:226–230.
6. Datta V, Chang A, Mackay S, et al. The relationship between motion analysis and surgical technical assessments. *Am J Surg.* 2002;184:70–73.
7. Helmreich RL, Merritt AC, Wilhelm JA. The evolution of Crew Resource Management training in commercial aviation. *Int J Aviat Psychol.* 1999;9:19–32.
8. Holzman RS, Cooper JB, Gaba DM, et al. Anesthesia crisis resource management: real-life simulation training in operating room crises. *J Clin Anesth.* 1995;7:675–687.
9. Chopra V, Bovill JG, Spierdijk J, et al. Reported significant observations during anaesthesia: a prospective analysis over an 18-month period. *Br J Anaesth.* 1992;68:13–17.
10. 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.
11. Pandey V, Moorthy K, Wolfe J, et al. Procedural rating scales increase objectivity in surgical assessment. *Br J Surg.* 2003;90(suppl 1):14.
12. Operative Competence ('Op Comp') forms. www.jchst.org, 2003.
13. Healey AN, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. *Qual Saf Health Care.* 2004;13 (suppl 1):i33–i40.
14. Calland JF, Guerlain S, Adams RB, et al. A systems approach to surgical safety. *Surg Endosc.* 2002;16:1005–1014.
15. Lossing AG, Hattwell EM, Gilas T, et al. A technical-skills course for 1st-year residents in general surgery: a descriptive study. *Can J Surg.* 1992;35:536–540.

16. Anastakis DJ, Regehr G, Reznick RK, et al. Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg*. 1999;177:167–170.
17. Datta V, Bann S, Beard J, et al. Comparison of bench test evaluations of surgical skill with live animal operating performance assessments. *Br J Surg*. 2002;89(suppl 1):92.
18. Beard J, Rochester J, Thomas W. Assessment of basic surgical skills in the laboratory. *Br J Surg*. 2003;90(suppl 1):90.
19. Munz Y, Moorthy K, Bann S, et al. Ceiling effect in technical skills of surgical residents. *Am J Surg*. 2004;188:294–300.
20. Fletcher G, Flin R, McGeorge P, et al. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth*. 2003;90:580–588.
21. Gaba DM, Howard SK, Flanagan B, et al. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology*. 1998;89:8–18.
22. Avermate van JAG, Kruijsen EAC. NOTECHS: the evaluation of non-technical skills of multi-pilot air crew in relation to the JAR-FCL requirements. *Research Report for the European Commission* [DGVII NLR-CR-98443]. 1998; National Aerospace Laboratory, Amsterdam.
23. Gaba DM, Howard SK, Small SD. Situation awareness in anesthesiology. *Hum Factors*. 1995;37:20–31.
24. Gaba DM. Human error in dynamic medical domains. In: Bogner MS, ed. *Human Error in Medicine*. Hillsdale, NJ: Erlbaum, 1994:197–224.
25. Helmreich RL, Merritt AC. Safety and error management: the role of Crew Resource Management. In: Hayward BJ, Lowe AR, eds. *Aviation Resource Management*. Aldershot, UK: Ashgate, 2000:107–119.
26. Xiao Y, Mackenzie CF. Collaboration in complex medical systems: collaborative crew performance in complex operational systems. NATO Human Factors and Medicine Symposium, April 20–24, 1998. Neuilly sur Seire: NATO, 1998:4-1–4-10.
27. Xiao Y, Hunter WA, Mackenzie CF, et al. Task complexity in emergency medical care and its implications for team coordination. LOTAS Group: Level One Trauma Anesthesia Simulation. *Hum Factors*. 1996;38:636–645.
28. Helmreich RL, Schaefer HG. Team performance in the operating theatre. In: Bogner MS, ed. *Human Error in Medicine*. Hillside, NJ: Erlbaum, 1994:225–253.
29. Sexton JB, Helmreich R. Communication and teamwork in the surgical operating room. Panel Presentation given at the 2000 Aerospace Medical Association conference in Houston. *University of Texas at Austin Human Factors Research Project* [Technical Report 00–04]. Houston, 2000.
30. Sexton JB, Thomas EJ, Helmreich RL. Error, stress and teamwork in medicine and aviation: a cross-sectional study. *Chirurg*. 2000;71 (suppl):42.