The American Journal of Surgery (2013) 205, 745–751

The American Journal of Surgery®

Surgical Education

The Surgeons' Leadership Inventory (SLI): a taxonomy and rating system for surgeons' intraoperative leadership skills

Sarah Henrickson Parker, Ph.D.,[a,b,*] Rhona Flin, Ph.D.,[a] Aileen McKinley, F.R.C.S.Ed.,[c] Steven Yule, Ph.D.[a,d]

[a]*School of Psychology, University of Aberdeen, Aberdeen, UK;* [b]*National Center for Human Factors Engineering in Healthcare, MedStar Institute for Innovation, MedStar Health, Washington, DC, USA;* [c]*Department of Surgery, Aberdeen Royal Infirmary, Aberdeen, UK;* [d]*Department of Surgery/STRATUS Center for Medical Simulation, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA*

KEYWORDS:
Leadership;
Nontechnical skills;
Safety;
Surgeon;
Operating room

Abstract

BACKGROUND: Surgeons must demonstrate leadership to optimize performance and maximize patient safety in the operating room, but no behavior rating tool is available to measure leadership.

METHODS: Ten focus groups with members of the operating room team discussed surgeons' intraoperative leadership. Surgeons' leadership behaviors were extracted and used to finalize the Surgeons' Leadership Inventory (SLI), which was checked by surgeons ($n = 6$) for accuracy and face validity. The SLI was used to code video recordings ($n = 5$) of operations to test reliability.

RESULTS: Eight elements of surgeons' leadership were included in the SLI: (1) maintaining standards, (2) managing resources, (3) making decisions, (4) directing, (5) training, (6) supporting others, (7) communicating, and (8) coping with pressure. Interrater reliability to code videos of surgeons' behaviors while operating using this tool was acceptable ($\kappa = .70$).

CONCLUSIONS: The SLI is empirically grounded in focus group data and both the leadership and surgical literature. The interrater reliability of the system was acceptable. The inventory could be used for rating surgeons' leadership in the operating room for research or as a basis for postoperative feedback on performance.

Inadequate leadership was a contributing factor in 50% of sentinel (adverse) events in acute health care in 2006.[1] Leadership has been studied in critical care teams[2] and trauma resuscitation teams[3,4] but has yet to be sufficiently examined in the intraoperative setting.[5] Professional bodies recognize leadership as a core competence for operating surgeons,[6,7] and leadership is highly valued among surgeons and operating team members.[8] Although leadership is considered an essential surgical competence,[9,10] empiric evidence on the specific leadership behaviors and skills required of surgeons during operations is lacking.

The Accreditation Council for Graduate Medical Education in the United States considers nontechnical skills (NTS) as core competencies in medical training.[11] NTS are defined as a combination of cognitive and social skills that complement technical skills and knowledge to create safe and efficient performance.[12] Cognitive skills typically include decision making, situation awareness, and task management. Social skills include leadership, teamwork, and com-

This work was funded by the Scottish Funding Council Strategic Research Development Grant to the Scottish Patient Safety Research Network. The authors have no conflicts of interest to declare.

* Corresponding author. Tel.: 202-244-9816; fax: 202-537-6052.
E-mail address: sarah.henrickson.parker@gmail.com

Manuscript received October 14, 2011; revised manuscript February 27, 2012

0002-9610/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amjsurg.2012.02.020

Table 1 Focus group participant demographics

Discipline	Participant group	n per group	Experience level per group	Length (h:min:s)
Surgeon	1	3	3 to 7 y postconsultancy	01:08:51
	2	9	1 to >15 y postconsultancy	00:45:25
	3	4	13 to 30 y postconsultancy	00:51:32
Trainee	1	14	ST 1–6, SPR 4, 5	01:04:07
	2	4	SPR 3–5	00:58:22
Nurse	1	8	>5 y experience in theater	00:44:11
	2	4	3 to 25 y experience in theater	00:47:45
	3	3	3 to >10 y experience in theater	00:31:50
Anesthetist	1	4	13 to 32 y postconsultancy	01:07:05
	2	11	2.5 to 15 y postconsultancy	00:49:41

SpR = specialist registrar; ST = surgical trainee.

munication. Currently, NTS are not widely taught in health care[13] but can be trained through on-the-job methods, lectures about teamwork, or simulation-based training.[14]

To effectively evaluate NTS, rating tools must be developed with clearly defined, observable elements.[15] A combination of task analysis methods such as observations, interviews, or subject matter expert review can be used to ensure that the NTS taxonomy is context specific and has content validity. These methods clarify exemplar behaviors to be associated with each element. Thus, the behaviors can be observed and the skills rated by a trained assessor.

Taxonomies and related tools to train and evaluate NTS have been developed in many industries, such as aviation,[16] anesthesia,[17] and surgery.[18] Once developed, these systems were subject to systematic experimental and practical evaluation, ensuring that the identified skills and associated behaviors were observable and could be reliably identified and rated. Of particular importance for applied researchers is that the tools are usable within the work domain and have face validity with practitioners.

Task analyses conducted to develop the Non-Technical Skills for Surgeons taxonomy revealed that expert surgeons believed leadership to be important for safety and efficiency in the operating room (OR).[18,19] Although Non-Technical Skills for Surgeons and other intraoperative NTS rating tools, such as the Oxford Non-Technical Skills[20] and Revised Non-Technical Skills[21] tools, have examined surgeons' leadership as part of a larger skill set, these tools define leadership in different ways and suggest different component elements of leadership, and none is sufficiently detailed for an in-depth analysis of surgical leadership. To our knowledge, a sufficiently fine-grained taxonomy of surgeons' leadership skills has not been developed, and consequently there is no evidence-based tool for identifying and classifying surgeons' intraoperative leadership behaviors.[5]

The aim of this study was to identify elements of surgeons' intraoperative leadership and then finalize and test the reliability of a prototype rating tool, the Surgeons' Leadership Inventory (SLI). In this report, we describe 2 studies. Study 1 was a focus group study conducted with surgeons, nurses, anesthetists, and trainee surgeons to identify surgeons' leadership behaviors in the intraoperative setting and resulted in a preliminary taxonomy, the SLI. In study 2, the interrater reliability of this taxonomy was tested using video recordings of live operations.

Study 1: focus group

Methods

This study was approved by the University of Aberdeen School of Psychology and North of Scotland Research Ethics Committees.

Ten single-discipline focus groups were arranged (see Table 1 for demographic information). There were no selection criteria for participation in the focus groups; participation was open to the entire OR team, and everyone who volunteered to participate was included in the study.

Single-discipline groups were chosen to engage the maximum number of staff members, while minimizing possible bias from interactions across disciplines, and to enable participants to enter the focus group sessions even if their entire operating teams were not present or did not want to participate. Ideally, focus groups should include between 4 and 8 people.[22] However, on 3 occasions, >8 participants volunteered, and it was decided to include all participants who showed an interest in taking part.

The purpose of the focus groups was to identify leadership behaviors of surgeons in the intraoperative setting. Although the surgeon may not be the sole leader of the surgical team, he or she retains responsibility for the outcome of the operation and is the de facto leader of the intraoperative team. To achieve this goal, a question schedule was developed on the basis of the surgical and psychological literature on surgeons' leadership.

The focus groups took place over 2 months (November and December 2009) and were facilitated by 1 psychologist (S.H.P.). All groups lasted between 30 minutes and 1 hour and were audio recorded. After completion of all 10 focus groups, the digital audio files were transcribed verbatim for analysis.

Statements on surgeons' leadership behaviors were extracted from all focus groups' transcripts and were assigned short summary codes intended to simplify the verbatim content of the behaviors. The behaviors were assigned to a preliminary taxonomy of surgeons' intraoperative leadership, which had been developed on the basis of a review of the literature[5] and an observation study.[23] Any behaviors that did not fit into the preliminary taxonomy were listed separately and were discussed by a group of subject matter experts.

The development of the taxonomy was based on 3 essential stages.[24] First, the target behavioral domain must be clearly defined, which in this case was the leadership behaviors of surgeons in intraoperative teams. The second step was to ensure that the observable outcomes were related to the behaviors. The outcome of interest was patient safety and team performance. Thus, all behaviors to be included in this taxonomy were expected to influence patient safety and team performance, although this relationship may not be direct. Testing of this relationship was outside the scope of this report, but according to expert review and the leadership literature, this relationship would be expected. The third step was to ensure that the taxonomy was internally and externally valid. Using the subject matter experts in both surgery and industrial psychology, we provide an initial examination of the taxonomy's validity. Additionally, the reliability of using the taxonomy was tested in study 2.

Results

It was clear from this exercise that some behaviors mentioned by the focus groups revealed leadership behaviors that had not been previously identified for surgeons, and therefore the taxonomy was revised. See Table 2 for the preliminary and final leadership taxonomies.

Definitions for each behavior were further clarified by the 3 psychologists (S.H.P., S.Y., R.F.), and the revised taxonomy and definitions were given to 6 subject matter experts (consultant or attending surgeons) familiar with NTS for review to establish face validity. Other surgical skills tests have used face validity as an initial estimate of validity,[25,26] and further testing of construct or criterion-related validity is necessary to determine the validity of a test.[27] Minor wording changes to the definitions were integrated, but no major changes were suggested. The final taxonomy, the SLI, is shown in Table 3.

Study 2: testing the taxonomy

Methods

To test the SLI for reliability, a series of videos of live surgical operations were analyzed. The study was approved by National Research Ethics Service of the National Health Service and the University of Aberdeen School of Psychology ethical review board.

The SLI (Table 3) was used by 2 psychologists to independently code video recordings of live operations using Studiocode version 2.5.49 (Studiocode Business Group, Sydney, Australia). Coders were trained during two 2-hour training sessions. The first centered on learning how to use the software, and the second session focused on learning the SLI definitions and behaviors included for each element.

To test interrater reliability, videos (n = 5) ranging from 55 minutes to 2 hours 30 minutes in length were coded, showing each operation from patient entry into the OR until placement of the last suture. Operations selected for analysis included 2 right hemiarch replacements, 1 inguinal hernia repair, 1 laparoscopic cholecystectomy, and 1 carotid endarterectomy.

Studiocode enabled the coders to watch each video in its entirety but with the facility to pause, rewind, and review any segment. A graphical representation of the elements in the SLI was created and added to the program, which is visible in Figure 1. The software provides a digital video, a timeline, and a clickable graphic of each possible element in the SLI (Fig. 1). Coders were able to categorize each leadership behavior by clicking on the element "button" in the on-screen graphic. The behavior was coded and time-stamped in the timeline. Coding of the 5 videos took place

Table 2 Development of the SLI

Preliminary taxonomy*	Modification	SLI
Maintaining standards	No change	Maintaining standards
Managing tasks	Behaviors included in "directing"	—
Making decisions	No change	Making decisions
Managing resources	No change	Managing resources
Directing and enabling	"Enabling" removed	Directing
Guiding and supporting	Split into "training" and "supporting others"	Training
		Supporting others
Communicating and coordinating	"Coordinating" removed	Communicating
	Newly added: "coping with pressure"	Coping with pressure

*Identified in Henrickson Parker et al.[5,23]

Table 3 The SLI elements and definitions

Element	Definition
Maintaining standards	Supporting safety and quality by adhering to acceptable principles of surgery, following codes of good clinical practice, and enforcing theater procedures and protocols by consistently demonstrating appropriate behaviors (ie, asking for help when required).
Making decisions	Seeking out appropriate information and generating alternative possibilities or courses of action, synthesizing the information, choosing a solution to a problem, and letting all relevant personnel know the chosen option. Making an informed prompt judgment on the basis of information, clinical situation, and risk and continually reviewing its suitability in light of changes in the patient's condition.
Managing resources	Assigning resources (people and equipment) depending on the situation or context; delegating tasks appropriately to team members, and ensuring the team has what it needs to accomplish the task.
Directing	Clearly stating expectations regarding accomplishment of task goals; giving clear instructions; using authority where required; demonstrating confidence in both leadership and technical ability.
Training	Instructing and coaching team members according to goals of the task; modifying own behavior according to team's educational needs; identifying and maximizing educational opportunities.
Supporting others	Judging the capabilities of team members; offering assistance where appropriate; establishing a rapport with team members and actively encouraging them to speak up.
Communicating	Giving and receiving information in a timely manner to aid establishment of a shared understanding among team members. Speaking appropriately for the situation. Asking for input from team members.
Coping with pressure	Showing flexibility and changing plans if required to cope with changing circumstances to ensure that goals are met; anticipating possible complications and communicating them to staff; adopting a forceful manner if appropriate without undermining the role of other team members.

over 1 week, after which interrater reliability was calculated using Cohen's κ, with criteria of $>.6$ as acceptable, $>.7$ as good, and $>.8$ as excellent.[28] The κ coefficient was chosen for interrater reliability because it assesses the degree to which different raters or observers give consistent estimates to the same phenomenon. For this study, it was considered important that 2 different raters could identify and code the same leadership behaviors at an acceptable level of reliability.

Results

Interrater reliability using the SLI across the 4 videos was $\kappa = .70$, which is considered good.[28] Directing and training behaviors were the most challenging to reliably code and were most often interchanged by the 2 coders.

Comments

The SLI was designed to provide an evidence-based skill set for training and to also allow structured observation and assessment of surgeons' leadership in the intraoperative setting. The 8 elements were empirically developed through an iterative process including a literature search, observations of surgeons' leadership, and focus groups with different groups of clinical personnel. After minimal training, 2 coders were able to use the taxonomy to identify and categorize leadership behaviors in the videos of live operations at an acceptably reliable level. This study has taken an additional step beyond the available NTS taxonomies for surgeons discussed above, by focusing on 1 particular category of skill in detail. The SLI is advantageous because it has been iteratively developed using subject matter experts in both psychology and surgery.

In industrial models of team leadership, goal setting and clarification, as well as team maintenance behaviors, such as developing a positive climate within the team, solving interpersonal problems, or developing group cohesion, are important for effective leadership.[29] However, from the focus groups, surgeons do not seem to engage in team maintenance behaviors, as would be expected according to the leadership literature. This may be because the goal of the task is discussed preoperatively or is implicit (to get the patient through surgery safely) and therefore is not specifically discussed in the intraoperative phase. Or it could be because the technical tasks of the operating team differ in such a way that there is not a shared task that unites the team other than to, again, get the patient through surgery safely. Surgeons may not see it as part of their leadership role to always ensure that the team is adequately cared for ("maintained"), because the team is often ad hoc, and the surgeon is only in a team supervisory role for the intraoperative phase of the operation, which is a relatively short period. In the focus groups, surgeons stated that making sure the team is "on board" and "present" were important but did not discuss managing conflict or building commitment between team members. Surgical teams who remain working together in small hospitals or in very specialized areas may show different team maintenance behaviors.

Many other NTS were alluded to when gathering the data to develop the leadership taxonomy, as may have been

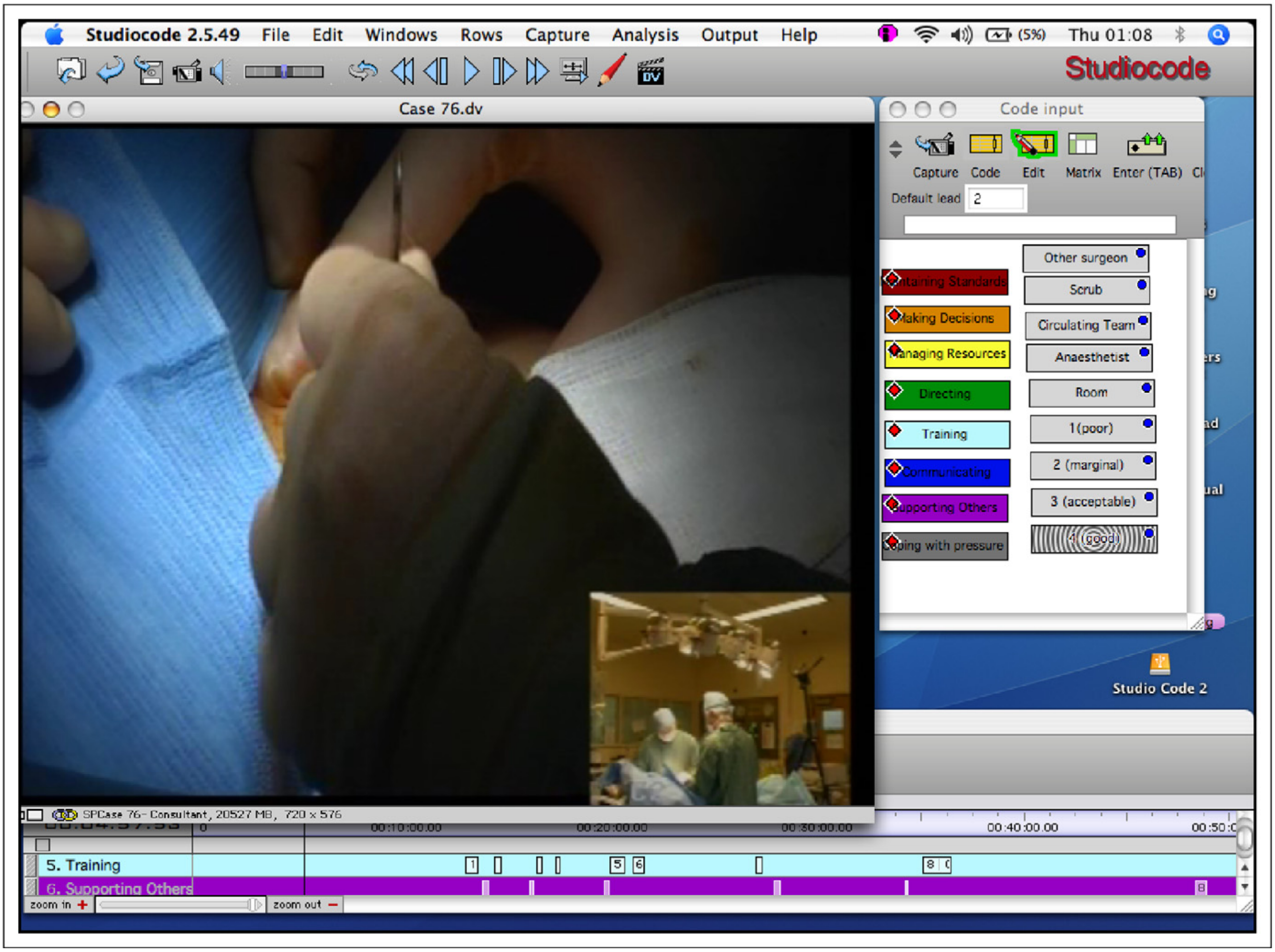

Figure 1 Screenshot of the Studiocode interface.

expected given the interrelatedness of critical NTS in practice. For example, surgical decision making has been examined as an NTS itself,[30] separate from leadership. However, in this study, operating team members frequently discussed the surgeon's ability to make decisions as imperative to effective intraoperative leadership. In their model of team leadership, Hughes et al[31] include decision making as an integral team leadership function, essential for teams to complete their task.

Because most NTS are observed through communication, it is difficult to determine if communication should be classified as an independent NTS in and of itself. For example, in the Anaesthetists' Non-Technical Skills framework, communication is not explicitly mentioned, but "exchanging information" is an element of teamwork.[17] Similarly, in the Non-Technical Skills for Surgeons system, communication is grouped with teamwork as a category.[32] However, in this study, communication was viewed by the focus groups as an element of leadership and was therefore included as a discrete element. Most models of leadership (for an overview, see Fleishman et al[33]), at both the team and the executive levels, include the ability to communicate with followers and influence them to accomplish a goal.

In phase 2 of the present study, the reliability of the SLI was tested using video recordings of live operations. Although the overall interrater reliability of the taxonomy was good, some elements were less reliably coded than others, specifically directing and training. Within academic hospital settings, the issue of training and directing becomes especially prominent, because consultant surgeons must balance the need for high patient care standards with the need to offer trainees opportunities for independent practice.[34] These 2 elements could be more reliably identified in the future if further distinction is made between them in coder training. It is also possible that because all the videos were recorded in a teaching hospital, training behaviors were much more common than they would be in a nonteaching hospital.

As is common in applied research, there are some limitations to this study, which we did our best to minimize. Only staff members from large teaching hospitals were included in this study, so the results may need to be further tested in smaller hospitals and nonteaching hospitals. The 8 elements for surgeons' intraoperative leadership would probably remain the same, with the exception of training for nonteaching hospitals. This also highlights the issue of the generalizability of the results given the small sample size. In an attempt to capture the views of a broad range of participants, surgeons from various disciplines and geographic locations within Scotland, as well as nonsurgeons, were included. However, the SLI needs to be tested and validated in a number of settings, such as simulations of live operations or multiple types of surgery (eg, those disciplines that were not included in the development phase), before we are confident that it can be considered a globally applicable model of surgeons' intraoperative leadership.

Surgeons' leadership could differ depending on the type of operation, because there are different requirements depending on the clinical task. For example, during cardiac surgery, the perfusion aspect of the case and larger team might add an additional level of leadership requirements. Future research should include a larger variety of disciplines to account for this. Researcher bias could also have influenced the results, because the researcher was also the facilitator of the focus groups. However, because the focus groups were based on a predetermined structure, this bias was minimized as much as possible. An additional limitation is that only surgeons were recruited to do a final review of the taxonomy, although it had been developed with experts from different disciplines in the OR. Because surgeons' leadership was the focus of the tool, it seemed logical for the subject matter experts to be involved in the final review. Future research could include other disciplines in a review of the SLI to further refine the definitions of each element. Also, reliability tests were conducted between 2 psychologists and did not include any clinical experts. Because the interest of this study was nonclinical (ie, leadership), it was decided that clinical knowledge was not deemed necessary to identify leadership behaviors for the purposes of reliability testing. However, future studies should test the reliability of the taxonomy using surgeons as raters, because their clinical perspective may distinguish their identification of leadership behaviors compared with psychologists. It would be especially important to establish this if the SLI were to be used for feedback by surgeons after real cases in the OR. Finally, although the κ coefficient for interrater reliability using the taxonomy to rate the videos showed substantial agreement, it is expected that this would improve further as raters receive more training in and practice using the taxonomy.

Conclusions

We describe the task analytic methods for development, and preliminary reliability testing, of the SLI. The SLI was found to be reliable when coding videos of live operations. The next phase of research should focus on the quality or effectiveness of these behaviors within the OR. It is also important to examine the usability and reliability of this taxonomy with clinical experts. In the future, this taxonomy could be used by appropriately trained observers for assessing surgeons' leadership in the intraoperative setting and as a tool for developing leadership training programs in surgery.

References

1. Joint Commission on Accreditation of Healthcare Organizations. Update: sentinel event statistics. Jt Comm Perspect 2006;26:14–5.
2. Kunzle B, Kolbe M, Grote G. Ensuring patient safety through effective leadership behaviour: a literature review. Saf Sci 2010;48:1–17.
3. Klein KJ, Ziegert JC, Knight AP, et al. Dynamic delegation: shared, hierarchical, and deindividualized leadership in extreme action teams. Adm Sci Q 2006;51:590–621.
4. Xiao Y, Seagull FJ, Mackenzie CF, et al. Adaptive leadership in trauma resuscitation teams: a grounded theory approach to video analysis. Cog Technol Work 2004;6:158–64.
5. Henrickson Parker SE, Yule S, Flin R, et al. Towards a model of surgeons' leadership in the operating room. BMJ Qual Saf 2011;20: 570–9.
6. Dickinson I, Watters D, Graham I, et al. Guide to the assessment of competence and performance in practising surgeons. ANZ J Surg 2009;79:198–204.
7. University of Toronto. CanMEDS framework. Available at: http://www.deptmedicine.utoronto.ca/CanMEDS.htm. Accessed: July 30, 2012.
8. American College of Surgeons. Frequently asked questions about surgical residency 2001. Available at: http://www.facs.org/medicalstudents/answer6.html. Accessed May 21, 2009.
9. Giddings A, Williamson C. The leadership and management of surgical teams. London: Royal College of Surgeons of England; 2007.
10. Krummel TM, Salzberg AD, Williams DB, et al. Accountability in surgery. In: Manuel BM, Nora PF, eds. American College of Surgeons surgical patient safety: essential information for surgeons in today's environment. Chicago: American College of Surgeons; 2004.
11. Accreditation Council for Graduate Medical Education. Program director guide to the common program requirements. Chicago: Accreditation Council for Graduate Medical Education; 2008.
12. Flin R, O'Conner P, Crichton M. Safety at the sharp end. Hampshire, UK: Ashgate; 2008.
13. Flin R, Patey R. Improving patient safety through training in nontechnical skills. BMJ 2009;339:3595.
14. Rosen MA, Feldman M, Salas E, et al. Challenges and opportunities for applying human factors methods to the development of nontechnical skills in healthcare education. In: Duffy VG, ed. Advances in human factors and ergonomics in healthcare. London: CRC Press; 2011:1–10.
15. Klampfer B, Flin R, Helmreich RL, et al. Enhancing performance in high risk environments: recommendations for the use of behavioural markers. Available at: http://homepage.psy.utexas.edu/homepage/group/helmreichlab/Publications/pubfiles/Pub262.pdf. Accessed July 21, 2012.
16. Flin R, Martin L, Goeters K, et al. Development of the NOTECHS (Non-Technical Skills) system for rating pilots' CRM skills. Hum Factors Aerospace Saf 2003;3:95–117.
17. Fletcher G, Flin R, McGeorge M, et al. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. Br J Anesth 2003;90:580–8.
18. Yule S, Flin R, Maran N, et al. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. World J Surg 2008;32:548–56.
19. Yule S, Flin R, Paterson-Brown S, et al. Non-technical skills for surgeons in the operating room: a review of the literature. Surgery 2006;139:140–9.
20. Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS system: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. BMJ Qual Saf 2009;18:104–8.

21. Sevdalis N, Davis R, Koutantji M, et al. Reliability of a revised NOTECHS scale for use in surgical teams. Am J Surg 2008;196: 184–90.
22. Kitzinger J. Qualitative research: introducing focus groups. BMJ 1995; 311:299.
23. Henrickson Parker SE, Yule S, Flin R, et al. Surgeons' leadership in the operating room: an observational study. Am J Surg 2012;204:347–54.
24. Fleishman EA, Quaintance MK. Taxonomies of human performance: the description of human tasks. Orlando, FL: Academic Press; 1984.
25. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. Ann Surg 2004;240:518–28.
26. Moorthy K, Munz Y, Sarker SK, et al. Objective assessment of technical skills in surgery. BMJ 2003;327:1032–7.
27. Korndorfeer JR, Kasten SJ, Downing SM. A call for the utilization of consensus standards in the surgical education literature. Am J Surg 2010;199:99–104.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
29. Kozlowski SWJ, Gully SM, Salas E, et al. Team leadership and development: theory, principles, and guidelines for training leaders and teams. In: Beyerlein M, ed. Advances in interdisciplinary studies of work teams: team leadership, vol 3. New York: Elsevier Science/ JAI Press; 1996:253–91.
30. Pauley K, Flin R, Yule S, et al. Surgeons' intraoperative decision making and risk management. Am J Surg 2011;202:375–81.
31. Hughes R, Ginnett R, Curphey G. Leadership: enhancing the lessons of experience. Homewood, IL: Irwin; 1993.
32. Yule S, Flin R, Paterson-Brown S, et al. Development of a rating system for surgeons' non-technical skills. Med Educ 2006;40:1098–104.
33. Fleishman EA, Mumford MD, Zaccaro SJ, et al. Taxonomic efforts in the description of leader behavior: a synthesis and functional interpretation. Leadership Q 1991;2:245–87.
34. Moulton C, Regehr G, Lingard L, et al. Operating from the other side of the table: control dynamics and the surgeon educator. J Am Coll Surg 2010;210:79–86.