

ORIGINAL ARTICLE

A literature review of educational feedback in the operating room: plastic surgery residents' perception of feedback from the O-SCORE

Yaeesh Sardiwalla,^{a,*,†} Nadim Joukhadar^{a,b} and David Tang^{a,b}

^aFaculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada; ^bDivision of Plastic and Reconstructive Surgery, Dalhousie University, Halifax, Nova Scotia, Canada

*Corresponding author at: Clinical Research Center – Room C125, 5849 University Avenue, Halifax, B3H 4R2, Canada.
Email: yaeesh@dal.ca

†Present address: Division of Plastic and Reconstructive Surgery, McMaster University, Hamilton, Ontario, Canada

Date accepted for publication: 19 April 2019

Abstract

Introduction: Surgical teaching in the operating room represents a unique and distinct aspect of medical education. Traditionally, an apprenticeship model of learning and feedback has been used. Recent changes in residency education have demanded a reevaluation of this model, resulting in the pursuit of improving educational feedback in the operating room through structured feedback. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) evaluates surgical skills and was recently introduced to the Plastic Surgery Training Program in addition to the traditional In-Training Evaluation Report (ITER). **Methods:** Plastic surgery residents (postgraduate years 2–5, $N = 9$) who were exposed to the transition between using ITER and O-SCORE evaluations were recruited. A grounded theory approach was used to analyze a semi-structured focus group with the residents. In addition, a 5-minute survey contrasting ITER and O-SCORE methods was distributed. **Results:** Residents suggested the O-SCORE has led to more opportune feedback, more comprehensive discussion of surgical procedures, and improved progress tracking compared with traditional methods. Although there is a role for reactionary feedback, residents unanimously agreed that the O-SCORE heralded an improvement in their learning. **Conclusion:** The introduction of the O-SCORE to the Plastic Surgery Program has complemented traditional feedback. Even though the O-SCORE adds to the evaluation burden, it may have an important role in the assessment structure of surgical residency training programs.

Keywords: medical education; plastic surgery; teaching; O-SCORE; ITER; feedback

Introduction

Surgical teaching in the operating room represents a unique and distinct aspect of medical education. Effective surgical instruction should facilitate the development of complex technical skills but also facilitate the development of a learner's preoperative planning and dynamic intra-operative judgement. Traditionally, an apprenticeship model of learning, observing, and participating in surgery has been utilized.¹ Changes in residency education including work-hour restrictions that reduce opportunities to gain graduated operating independence demand the reevaluation of the apprenticeship model.^{2–4} In addition, pioneers in surgical education have recognized that improving educational feedback can maximize perioperative learning for trainees.

Previous studies describe that traditional feedback in the operating room consisted primarily of the attending surgeon commanding, questioning, and explaining instrumental interactions in response to resident error.^{5,6} It was rare for a resident to solicit this feedback or to be satisfied with the type of operative assessment they received.⁷ A structured debrief was only used in 46% of perioperative interactions.⁸ These practices should be contrasted with what residents and attending staff describe as effective. Components highlighted include deconstructing complex tasks in specific steps that enable graduated responsibility, allowing the resident to feel the pathology, and involving them in all aspects of surgical planning.^{9–11} In addition, a study that involved a multidisciplinary perspective

(surgeons, residents, anesthesiologists, and nurses) identified that creating a learning environment through a structured debriefing where surgical performance can be analyzed critically in a safe environment was important to learning.^{12,13}

There have been several structured evaluation and feedback tools proposed for such debriefings in the perioperative environment.^{14–16} These tools incorporate elements of effective feedback, including preoperative planning, technical performance during a case, decision making at key steps of the procedure, independence, and overall patient management. The use of these tools resulted in improved quality of feedback and comprehensiveness as perceived by trainees and staff.¹⁵ In addition, the operative performance by residents appears to benefit as a direct result of more descriptive feedback.¹⁷

Simulation has served as an effective tool to assess residents' competency in medical education.^{18,19} There is evidence reporting that an important aspect of simulation is the use of standardized feedback, or "debriefing".^{20,21} We proposed that using standardized, competency-based assessment tools for in-training clinical scenarios would improve residents' feedback for technical skills from preceptors. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) is an example of such a standardized assessment tool that was purposefully designed to assess a trainee's competence to perform entire surgical procedures independently.¹⁶ The O-SCORE is a structured 11-item tool that utilizes entrustment anchors on a 5-point scale, offering a construct-aligned assessment tool to rate a trainee's operative skills.²² The O-SCORE has also been shown to successfully discriminate surgical competency between junior and senior residents.^{16,23} However, there is a paucity of evidence regarding residents' experience with this evaluation tool.

The Plastic Surgery Training Program recently introduced the O-SCORE as a work-based assessment tool, adding to the existing evaluation of technical skills covered in the traditional In-Training Evaluation Report (ITER). A minimum of three O-SCORE evaluations was requested per 3-month rotation. We planned to use the period of transition to study more comprehensively what benefits and disadvantages residents perceive in each of the feedback tools. We hypothesized that the more structured feedback on technical performance provided by the O-SCORE would be more descriptive, and thus better valued by plastic surgery residents compared with the more traditional ITER approach.

Methods

Participant selection

Permission to conduct this study was obtained from the Research Ethics Board (REB no. 2017-4322). We invited

all plastic surgery residents at postgraduate year 2–5 levels to participate in the study. These were residents who were exposed to the addition of the O-SCORE for at least 1 year after the exclusive use of the ITER. Three O-SCORES for each rotation were required to be completed, with a collaborative approach to scheduling the assessment event between the resident and attending surgeon. Faculty were introduced to the O-SCORE evaluation format by the Program Director in a formal faculty development workshop.

Research design

The current transitory period at the Plastic Surgery Training Program was opportune to assess the impact of the O-SCORE with a retrospective cohort design. Residents who agreed to participate in the research took part in a semi-structured focus group that required them to reflect on their experience to date with the O-SCORE and contrast experiences to the ITER (see Fig. 1 for focus group guiding questions). Elements of reflexivity were used to guide the focus group with no preset outcome or expectation dictating the process. Focus group leaders followed these principles, were cognizant of their preconceptions and strove to lead knowledge production in the session. The discussion was emergent based on topics raised by residents.

The content disclosed in the focus group was recorded and transcribed anonymously. The focus group was led by two independent researchers not associated with the training program in order to avoid the introduction of confirmatory bias or power gradients. In addition, a survey (Fig. 2) was distributed asking residents to rate the descriptiveness, knowledge level identification, timeliness, clarity, and opportunity to discuss evaluations for both ITER and O-SCORE methods.

Analysis of data

A grounded theory strategy, the qualitative analysis method in which there are multiple stages of collecting, analyzing, and comparing the data, was used to process the focus group content. Phenomenology, the process of understanding a group or individuals' perspective free of bias through convergent and divergent experiences, was used to generate a narrative through themes about experiences with the O-SCORE and ITER from the residents' individual and group perspective. Reflexivity was used to guide the analysis, with two separate reviewers analyzing the transcript for personal biases, inappropriate assumptions, and dynamics of the interview that may have contributed to the messaging. Reflexivity refers to self-evaluation of researchers to explore patterns of both expected and unanticipated relationships in the content. Themes were assessed relative to those

1. During your plastic surgery rotations in the OR, how would you rate the quality of feedback given by staff?
 - a. How often do staff dedicate time for feedback (no other distractions)
 - i. How long do they give you?
 - b. Do you feel that training staff to use structured feedback/assessment tools is a good idea for enhancing your learning?
2. What makes useful feedback?
 - a. What's the best thing about the way you're getting feedback right now?
 - b. Can you tell me about a time when receiving feedback helped you learn more about a particular skill or competency?
 - c. What would you say to your staff regarding feedback so that it works better for you?
 - d. What's the worst thing about the way you're getting feedback now?
3. There are certain skills that need to develop during residency to become a competent plastic surgeon. Are there aspects of your performance that you wish were addressed more often?
 - a. Has there been a time that you did something particularly well or poorly but wasn't addressed? Why so?
4. Are there aspects of your performance that you wish were addressed more often?
 - a. Mentioned by colleagues:
 - i. Efficiency
 - ii. Feedback re: critical decision making in the face of complex situations
 - iii. Feedback outside of the OR – eg, floors, codes, traumas, emergencies
 - iv. Rare scenarios
 - v. Conflict and personality management
5. Would it be beneficial if an attending staff who is trained in the delivery of feedback train other staff in providing useful feedback for residents? What are your thoughts on that?
6. During the transition from using informal feedback and the ITER to using the O-SCORE, what changes in the quality of feedback have you noted? Has the change been for the better? What would you like changed?
7. General probe and open floor —does anyone else have anything to add to that? Does anyone want to share a similar or maybe different experience?

Figure 1. The guiding questions that were used to lead the semi-structured focus group. Discussion deviated from this script following areas residents wished to explore further.

previously established using a constant comparative analysis, a process in which concepts from the data are developed by coding and analyzing at the same time. Each theme was then explored utilizing verbatim quotes from the transcript to best represent residents' voices.

To analyze the survey results, median values and interquartile ranges (IQR) of Likert scores for the survey (1 = strongly agree and 5 = strongly disagree) were calculated. The survey dataset was non-normal, ordinal data so a non-parametric Wilcoxon signed-rank test was used to examine differences in feedback value between the two tools.

Results

Of ten eligible plastic surgery residents, nine participated in the focus group and six fully completed the survey. The results of the focus group have been subcategorized into four primary themes according to the analysis described in the Methods.

Traditional intraoperative feedback

Informal feedback from preceptors, although variable, was identified as important for a resident's learning, particularly intraoperatively. The quality and effectiveness of this feedback was stated to be highly faculty dependent and seemed to focus on items that trainees were doing incorrectly. Residents believed that the opportunity for this feedback was typically confined to the event or immediately after. They believed that the opportunity was lost if not discussed in the moment.

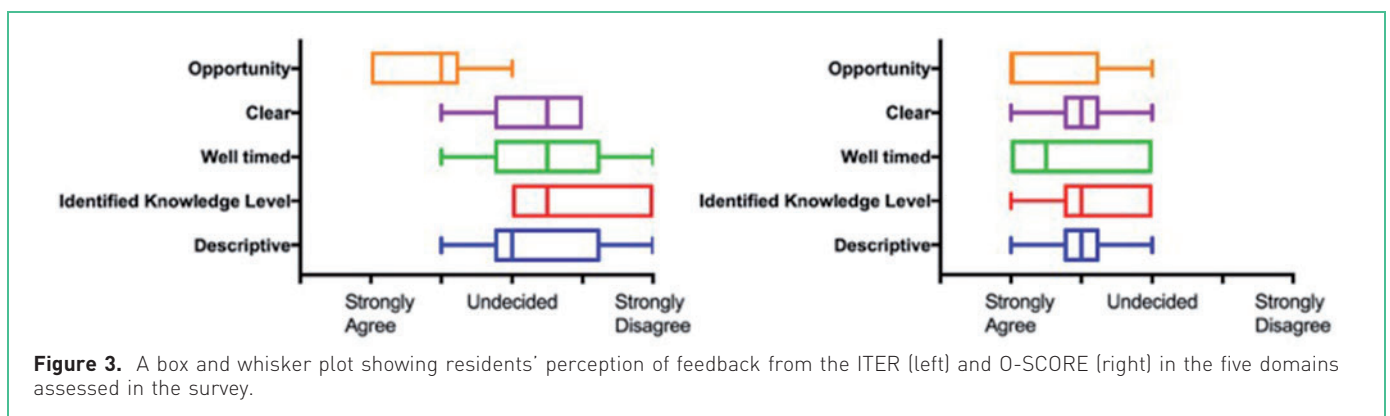
A lot of the feedback you get is reactive and it's usually something that is a good decision in the moment, or something that you're not executing effectively. Usually, you get more criticism with an idea so that you can correct it at the time that you're doing it. There's not a lot of proactive or active feedback throughout the case, unless residents specifically seek it out or ask a question.

I think it's very difficult to get feedback after a case because it's the end of the day and the next time you

Answer the following questions based on your perception of OSCORE evaluations

Question 1	5 Strongly Disagree	4 Disagree	3 Undecided	2 Agree	1 Strongly Agree
Feedback provided was detailed and descriptive enough to understand the competency of my technical skills in the OR	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Question 2	5 Strongly Disagree	4 Disagree	3 Undecided	2 Agree	1 Strongly Agree
The feedback provided appropriately identified my knowledge level regarding the operation and what needed further studying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Question 3	5 Strongly Disagree	4 Disagree	3 Undecided	2 Agree	1 Strongly Agree
The feedback was well timed and provided at a good frequency (not overwhelming or too little)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Question 4	5 Strongly Disagree	4 Disagree	3 Undecided	2 Agree	1 Strongly Agree
The feedback was clear and did not require clarification with my supervisor to understand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Question 5	5 Strongly Disagree	4 Disagree	3 Undecided	2 Agree	1 Strongly Agree
There was an opportunity to discuss the feedback with my supervisor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Final Question – Comparison ITER vs O-SCORE					
Which of the two evaluation tools is more effective in providing useful feedback on your technical skills?	ITER	O-SCORE			
	<input type="radio"/>	<input type="radio"/>			

Figure 2. An example of the survey distributed that asked residents to rate their perception of the O-SCORE. An identical form was used for the ITER.



see your staff you're doing something else. Unless there is dedicated time for feedback.

For me, it's about getting the reactionary feedback intraoperatively. This is helpful with intraoperative skills. Either changing your technique or watching the staff do something and trying to mimic that.

O-SCORE strengths

The introduction of the O-SCORE has allowed for noticeable improvements in both the quality and timing of feedback. By necessitating O-SCORE completion, an opportunity for discussion of a case was facilitated, permitting important aspects (preoperative, intraoperative, and postoperative) of the case to be defined more comprehensively. It allowed for feedback to be more specific according to entrustability anchors, enabling improvements in the same case to be tracked over time and therefore leading to more effective longitudinal learning. There was unanimous agreement that the O-SCORE provided improvements and complemented traditional feedback.

I think with the O-SCORE, the process of feedback is formalized. So basically, it forces you and staff to go through different elements of the surgery including pre- and postoperative plans. You get formal feedback for these higher processes. I feel if we didn't have this tool, the feedback wouldn't be as specific unless you purposefully sought it out with your staff.

I think the O-SCORE is definitely a step in the right direction. Since it has been introduced, the staff and I actually sit down and review it. You actually talk about the case more. But what I found was even more useful is when you pick a case that you're doing your O-SCORE for and decide this a few days before, it encourages playing a more active part in the planning.

O-SCORE weaknesses

A selection bias to use the O-SCORE only on certain cases for which the residents felt proficient to improve evaluations represents a shortcoming in its current implementation at the study institution. There was also a discussion of the right-shift tendency of overly generous evaluations when using these standardized tools that can undermine accuracy. The underutilization of the O-SCORE was contrasted to the growing evaluation burden demanded throughout residency.

Probably the worst thing about the O-SCORE and the reason we're all dreading it is the fact that if someone says you should do this after every single case. The fact is, they take quite a while to do, so if you have more boxes to tick and more sections it will take even longer.

It is important to find the balance of being useful and cumbersome.

One thing to note with the O-SCORE is that because we only have three that are mandatory per rotation, you can pick which cases to use. I'm going to pick a very standardized case which will get you all 5s so then you look great on your piece of paper.

Considerations for the future

The implementation of standardized feedback such as the O-SCORE offered an improvement in feedback that confers future promise. One suggestion to improve the process involves making debriefings a standardized aspect of a surgery day, which could naturally reduce the evaluation burden of the O-SCORE. Another proposal for improvement would involve having surgery specific (craniofacial, breast, hand, and microsurgery) forms, which may further increase the specificity and tracking that is able to be achieved.

I'm on a rotation now where we're doing them for 70% of cases and it's helpful and not onerous. It's efficient and is part of the day, but I feel if you were just doing three in three months, you sit down and have a fairly involved discussion, but if you're doing it more frequently, you just talk about what was most important about that case.

Plastic surgery is a vast speciality with a lot of variability for each procedure. Technical skills for each surgery are different. Hand surgery is not the same as breast surgery with different principles. Unless you're doing the same procedure, it's difficult to track your progress.

So, in an ideal world having more categories for the O-SCORE would be better. There could be more specific checkboxes depending on the procedure you're doing.

Survey analysis

All residents selected the O-SCORE as the preferred form of feedback. Median, IQR, and comparisons from non-parametric Wilcoxon signed-rank test for the survey are presented in [Table 1](#).

Discussion

With the recent introduction of Competence By Design (CBD) across Canadian residency programs,²⁴ the use of standardized evaluation tools has become more important and necessary. A typical method of resident evaluation of technical skills in training programs is the ITER. This document is often a summative assessment of a trainee's performance over a given time period or rotation, measuring

Table 1. A summary of the median [M], interquartile range (IQR), and corresponding Wilcoxon signed-rank test for survey response items

	Descriptive	Identified knowledge level	Well timed	Clear	Opportunity to discuss
ITER	3 (3–4)	4 (4–5)	5 (3–5)	4 (3–4)	2 (1–2)
O-SCORE	2 (2)	2 (2–3)	2 (1–3)	2 (2)	1 (1–2)
Wilcoxon signed-rank	$Z = -1.73, P = 0.083$	$Z = -0.45, P = 0.66$	$Z = -1.00, P = 0.32$	$Z = -0.82, P = 0.41$	$Z = -1.73, P = 0.083$

performance along a subjective Likert scale in order to provide an overall impression on aspects done well and those requiring improvement.²⁵ The ITER encourages discussion of feedback in a personal meeting between the resident and the supervisor after the rotation. However, the quality of feedback can be limited due to poor choice of rating scales, incomplete forms, and delayed completion.²⁶ Feedback is therefore non-standardized and may not be conducive to improving performance.²⁶ The O-SCORE can provide a fresh perspective in this domain as described in our study.

The plastic surgery resident cohort that participated in this study believed that the introduction of the O-SCORE to evaluate surgical skills in the OR has benefitted the quality and timing of the feedback received. The O-SCORE served as a useful tool to facilitate a discussion of multiple components of the surgery by providing a schematic to initiate a conversation with faculty. These improvements are demonstrated by the focus group themes discussed and the improvements seen through the survey.

Concerns evident in traditional feedback using ITERs and informal reactive feedback included the variability between faculty, conversations being rushed, no discussion of pre- and postoperative planning, and the solely reactionary nature of the content. Although residents certainly agreed that there is a role for reactionary feedback in surgical education, the O-SCORE provided important complementary information that was valuable to their learning. A highlight of the O-SCORE was the emphasis on comprehensive patient care during the surgical experience. Even though the focus remained on technical skills, the incorporation of surgical planning that included assessing suitability for surgery, deciding on an appropriate operative plan, and suitable appropriate follow-up were welcome additions. Furthermore, the O-SCORE allowed for a more objective method of tracking these competencies over the duration of training.

The study had several limitations, including the lack of a validated survey to assess the feedback tools. The results must be interpreted accordingly and be used to provide descriptive information regarding the nature of the

O-SCORE. The cohort studied represent a small sample of all surgical residents and care must be taken when generalizing perspectives shared. We do, however, believe sample size ($N = 9$) for the focus was adequate to achieve thematic saturation.²⁷ There is also an intrinsic selection bias in O-SCORE administration by residents to select for cases they are proficient at to obtain good evaluations.²⁸ Plastic surgery residents are exposed to a set of unique challenges due to the variety of operations encountered, including hand, craniofacial, breast, and microsurgery, which was a source of concern from residents regarding generalizability of the assessment tool. In addition, each supervising surgeon may have individual preferences for the same procedure, thus complicating reliability of feedback considerations.

With these considerations in mind, there may be a role for further studying and refining the O-SCORE in randomly selected cases from different surgical specialties. This could help mitigate some of the limitations mentioned and enable better tracking of performance and progress through different years of residency with a larger cohort. By making O-SCORE evaluations more routine as suggested in the focus group, they become integrated to workflow and less burdensome. Although the tool has a time and efficiency expense that needs to be further examined, residents ascribe a beneficial role to the utilization in surgical CBD residency programs. Introductory training for residency programs (faculty and residents) on how to best use the O-SCORE could help with future transitions and clarify the role of this feedback as part of monitoring a resident's progress in the world of CBD.

Conclusion

The introduction of the O-SCORE to the Plastic Surgery Program has complemented traditional feedback. Residents suggest that it has led to more opportune feedback, more comprehensive discussion of surgical procedures, and improved progress tracking. Although the tool has a time expense and some limitations, it may have an important role in the assessment structure of surgical residency training programs.

Acknowledgements

Matthew Ta provided guidance in designing this project based on a research project conducted in the Department of Anesthesia, Pain Management and Perioperative Medicine at Dalhousie.

Conflict of interest

None declared

References

1. Reznick RK, Macrae H. Changes in the wind. *N Engl J Med* 2006; 355: 2664–2669. <https://doi.org/10.1056/NEJMra054785>.
2. Ahmed N, Devitt KS, Keshet I, Spicer J, Imrie K, Feldman L, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Ann Surg* 2014; 259: 1041–1053. <https://doi.org/10.1097/SLA.0000000000000595>.
3. Condren AB, Divino CM. Effect of 2011 Accreditation Council for Graduate Medical Education duty-hour regulations on objective measures of surgical training. *J Surg Educ* 2015; 72: 855–861. <https://doi.org/10.1016/j.jsurg.2015.04.018>.
4. Bell RH, Biester TW, Tabuenca A, Rhodes RS, Cofer JB, Britt LD, et al. Operative experience of residents in US general surgery programs. *Ann Surg* 2009; 249: 719–724. <https://doi.org/10.1097/SLA.0b013e3181a38e59>.
5. Blom EM, Verdaasdonk EGG, Stassen LPS, Stassen HG, Wieringa PA, Dankelman J. Analysis of verbal communication during teaching in the operating room and the potentials for surgical training. *Surg Endosc* 2007; 21: 1560–1566. <https://doi.org/10.1007/s00464-006-9161-0>.
6. Roberts NK, Brenner MJ, Williams RG. Original communications capturing the teachable moment: a grounded theory study of verbal teaching interactions in the operating room. *Surgery* 2012; 151: 643–650. <https://doi.org/10.1016/j.surg.2011.12.011>.
7. Jensen AR, Wright AS, Kim S, Horvath KD, Calhoun KE. Educational feedback in the operating room: a gap between resident and faculty perceptions. *Am J Surg* 2012; 204: 248–255. <https://doi.org/10.1016/j.amjsurg.2011.08.019>.
8. Ahmed M, Sevdalis N, Vincent C, Arora S. Actual vs perceived performance debriefing in surgery: practice far from perfect. *Am J Surg* 2013; 205: 434–440. <https://doi.org/10.1016/j.amjsurg.2013.01.007>.
9. Pernar LIM, Ashley SW, Smink DS, Zinner MJ, Peyre SE. Master surgeons' operative teaching philosophies: a qualitative analysis of parallels to learning theory. *J Surg Educ* 2012; 69: 493–498. <https://doi.org/10.1016/j.jsurg.2012.02.002>.
10. Iwaszkiewicz M, Darosa DA, Risucci DA. Efforts to enhance operating room teaching. *J Surg Educ* 2008; 65: 436–440. <https://doi.org/10.1016/j.jsurg.2008.07.006>.
11. van der Houwen C, Boor K, Essed GGM, Boendermaker PM, Scherpbier AAJJA, Scheele F. Gynaecological surgical training in the operating room: an exploratory study. *Eur J Obstet Gynecol Reprod Biol* 2011; 154: 90–95. <https://doi.org/10.1016/j.ejogrb.2010.08.014>.
12. Snyder RA, Tarpley MJ, Tarpley JL, Davidson M. Teaching in the operating room: results of a national survey. *J Surg* 2012; 69: 643–649. <https://doi.org/10.1016/j.jsurg.2012.06.007>.
13. Ahmed M, Sevdalis N, Paige J, Paragi-Gururaja R, Nestel D, Arora S. Identifying best practice guidelines for debriefing in surgery: a tri-continental study. *Am J Surg* 2012; 203: 523–529. <https://doi.org/10.1016/j.amjsurg.2011.09.024>.
14. Niitsu H, Hirabayashi N, Yoshimitsu M, Mimura T, Taomoto J, Sugiyama Y, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today* 2013; 43: 271–275. <https://doi.org/10.1007/s00595-012-0313-7>.
15. Leung Y, Salfinger S, Mercer A. The positive impact of structured teaching in the operating room. *Aust N Z J Obstet Gynaecol* 2015; 55: 601–605. <https://doi.org/10.1111/ajo.12392>.
16. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med* 2012; 87: 1401–1407. <https://doi.org/10.1097/ACM.0b013e3182677805>.
17. Grantcharov TP, Schulze S, Kristiansen VB. The impact of objective assessment and constructive feedback on improvement of laparoscopic performance in the operating room. *Surg Endosc* 2007; 21: 2240–2243. <https://doi.org/10.1007/s00464-007-9356-z>.
18. Lorello GR, Cook DA, Johnson RL, Brydges R. Simulation-based training in anaesthesiology: a systematic review and meta-analysis. *Br J Anaesth* 2014; 112: 231–245. <https://doi.org/10.1093/bja/aet414>.
19. Blum RH, Boulet JR, Cooper JB, Muret-Wagstaff SL. Simulation-based assessment to identify critical gaps in safe anesthesia resident performance. *Anesthesiology* 2014; 120: 129–141. <https://doi.org/10.1097/ALN.0000000000000055>.
20. Rall M, Manser T, Howard SK. Key elements of debriefing for simulator training. *Eur J Anaesthesiol* 2000; 17: 516–518. <https://doi.org/10.1097/00003643-200008000-00011>.
21. Okuda Y, Bryson E, DeMaria S, Jacobson L, Quinones J, Shen B LA. The utility of simulation in medical education: what is the evidence? *Mt Sinai J Med* 2009; 76: 330–343. <https://doi.org/10.1002/msj.20127>.
22. Budden CR. Using the Ottawa Surgical Competency Operative Room Evaluation (O-SCORE) in a Canadian Plastic Surgery Program. Doctoral dissertation, University of Alberta; 2016.

23. MacEwan MJ, Dudek NL, Wood TJ, Gofton WT. Continued validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): use in the simulated environment. *Teach Learn Med* 2016; 28: 72–79. <https://doi.org/10.1080/10401334.2015.1107483>.
24. Maudsley R, Dagg P, Wilson J, Frank J, Amit M, Brundage M, et al. Competence by design. Reshaping Canadian Medical Education. Royal College of Physicians and Surgeons of Canada. 2014. Available from <http://www.royalcollege.ca/rcsite/educational-initiatives/cbd-medical-education-ebook-e> (accessed 26 April 2019).
25. Weller JM, Misur M, Nicolson S, Morris J, Ure S, Crossley J, et al. Can I leave the theatre? A key to more reliable workplace-based assessment. *Br J Anaesth* 2014; 112: 1083–1091. <https://doi.org/10.1093/bja/aeu052>.
26. Bindal N. Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Med Teach* 2011; 33: 919. <https://doi.org/10.3109/0142159X.2011.558140>.
27. Ando H, Cousins R, Young C. Achieving saturation in thematic analysis: development and refinement of a codebook. *Compr Psychol* 2014; 3: 03. <https://doi.org/10.2466/03.CP.3.4>.
28. Rose JS, Waibel BH, Schenarts PJ. Disparity between resident and faculty surgeons' perceptions of preoperative preparation, intraoperative teaching, and postoperative feedback. *J Surg Educ* 2011; 68: 459–464. <https://doi.org/10.1016/j.jsurg.2011.04.003>.