

Cognitive and Technical Skill Assessment in Surgical Education: a Changing Horizon

Ashley Vergis^{1,2} · Krista Hardy¹

Received: 19 October 2010 / Accepted: 19 October 2010 / Published online: 10 March 2017
© Association of Surgeons of India 2017

Abstract Assessment is an integral component of training and credentialing surgeons for practice. Traditional methods of cognitive and technical appraisal are well established but have clear shortcomings. This review outlines the components of the surgical care assessment model, identifies the deficits of current evaluation techniques, and discusses novel and emerging technologies that attempt to ameliorate this educational void.

Keywords Surgical education · Assessment · Cognitive skills · Technical skills

Introduction

In a simplistic sense, the purpose of assessment and certification is to enhance learning and to ensure that the candidate is competent in all aspects required of their profession [1]. To this end, the character of assessment in medical education has been dissected, evaluated, and refined for more than two decades. This interest has led to broader notions of what assessment should be doing than in the past [2].

However, surgical residencies in the USA, Canada, and Western Europe have traditionally incorporated a limited number of methods to evaluate their learners. These techniques focus on technical and cognitive abilities and are limited in their scope of assessment. This review outlines the

components of the surgical care assessment model, identifies the deficits of current evaluation techniques, and discusses novel and emerging technologies that attempt to ameliorate this educational void.

Components of Surgical Care Assessment Methods

In order to understand the breadth of methods available to assess cognitive and technical competence, the surgical educator must first understand the scope of surgical practice. For this strategy, a useful schema describes four components of surgical care: diagnosis, treatment plan, technical performance, and postoperative care [3].

Diagnosis Diagnostic ability is essential in all areas of clinical medicine. In both acute and elective settings, a surgeon must be able to elicit a clear and relevant history, perform and interpret a focused physical exam, and request diagnostic tests. The surgeon must then analyze all of the data to formulate a differential diagnosis. Assessment should therefore include some method of determining how well a candidate can gather information and use it to generate the possible causes of the patient's presentation.

Treatment Plan With information gathered and interpreted, the surgeon must formulate a plan of action. This requires an understanding of available options, their alternatives and the relative merits of each. When developing a plan, the surgeon must also recognize his or her own abilities, the capabilities of the institution and be able to communicate this information with the patient. The plan and the surgeon must also be flexible and adapt as more data becomes available or as the situation evolves. In this context, assessment must be able to examine how well a candidate is able to integrate their medical knowledge and diagnostic abilities into a workable realistic process for the patient.

✉ Ashley Vergis
ashleyvergis@hotmail.com

¹ Section of General Surgery, University of Manitoba, Winnipeg, MB, Canada

² St. Boniface General Hospital, Z3039-409 Tache Avenue, Winnipeg, MB R2H 2A6, Canada

Technical Performance Technical dexterity is often incorrectly the sole focus when considering a surgeon's ability to perform a procedure. While technical dexterity is important, intraoperative decision-making is paramount to safely and successfully execute any surgical procedure.

Dexterity refers to the psychomotor aspects required to perform the planned procedure. In surgery, this is more than the eloquent movements that are esthetically satisfying to the untrained eye. Rather, it includes subtle motor abilities such as delicate tissue handling, correct tissue apposition, and tying sutures under the correct tension.

Intraoperative decision-making refers to the cognitive processes surgeons use to navigate through a procedure. This is the product of a surgeon's knowledge and judgment. Knowledge allows the surgeon to recognize intraoperative events and predict their effect on the procedure. Judgment relates to the relative importance the surgeon places on that event. For example, if a small bowel serosal tear is encountered intraoperatively, knowledge will tell the surgeon of the possible effects of this injury (leakage, stricture, or indolent). The surgeon's judgment will then be used to determine its importance (small tear means inconsequential, large tear means serious consequences). Based on these factors, the surgeon will decide whether or not the tear needs to be repaired.

As described, technical ability is a function of multiple competencies. Assessment methodology must therefore be robust and inclusive of the dexterity and decision-making skills required of a surgeon.

Postoperative care is a complex period that includes multiple disciplines and services. In this stage of the treatment plan the patient interacts with, among others, the surgeon, nurses, physiotherapists, and dieticians to ensure overall care. In this phase, the surgeon must evaluate and re-evaluate the patient's progress. If complications arise, the surgeon must utilize diagnostic, planning and, if necessary, technical abilities to correct them. This final component of surgical care illustrates the intricacies of surgical practice in that several aspects may interact simultaneously. Therefore, any assessment method employed should be flexible enough to measure a range of competencies simultaneously.

Current Assessment Methods in Surgical Education

Current assessment methods in postgraduate surgical training focus on the mastery of cognitive and technical outcomes. Traditionally, summative assessment is employed at the end of training for certification [4]. Knowledge objectives are typically assessed through written tests and oral examinations and technical skills are evaluated by final in-training evaluation reports (ITERS).

Cognitive Assessment

Written examinations have been a staple of medical assessment for many decades. Formats typically used in surgery include subjective instruments such as essays, and objective instruments such as supply-item (short answer and "fill in the blank"), multiple-choice, and extended matching tests. Overall, the unifying advantage of the written examination is that questions, marking schemes, and process can be standardized.

In general, the essay format allows free form response so that a candidate may select, integrate, and evaluate relevant information at length. However, although this format is useful for assessing a student's organizational skills and thinking process, it is problematic in surgical education. First, due to item length and the time required for marking, the essay is limited in the scope of content that can be assessed. As well, even with standardized marking schemes, reliability is extremely difficult to ensure due to the inherent variability in markers perception of style and organization. For these reasons, the essay format is rarely utilized in surgical assessment.

Multiple choice, supply item, and extended matching formats are examples of objective tests. Common to all is that they involve performing a structured activity for which a limited kind of response is possible. Answer schemes are absolute and quite reliable. These tests are useful for assessing factual information, are relatively easy to develop, do not require extensive resources to mark, and can measure a variety of learning outcomes. The key disadvantages are that examined content tends to be limited in depth and clinical performance remains un-examined.

The oral examination has been ingrained in surgical evaluation for decades. This format allows some assessment of facts but also allows for integration of information [3]. Traditionally, the candidate is challenged with a clinical vista and the case unfolds based on examinee response. The strength of this technique is that, based on responses, examiners can explore content in varying depth as appropriate. Also, the process by which the candidate arrives at a conclusion may be analyzed. This is realistic for clinical medicine as an incorrect conclusion based on sound reasoning is justifiable. Specific to surgical specialties, this format is able to assess intraoperative decision-making.

The chief difficulties with this method relate to standardization, objectivity, and assessment of performance. As this method is generally a "free flow" interaction between candidate and assessor(s), it is difficult to standardize format. Additionally, the ultimate structure of the scenario is determined by complex interactions between candidate and examiners that involve subjective interpretations of questions and responses. Furthermore, the oral examination may be seen as an intimidating process and some candidates may be disadvantaged by being more intimidated than others [3]. Finally,

this format only assesses what the candidate says they would/can do which may not reflect actual performance ability.

Technical Assessment

Theoretically, the ongoing analysis of a trainee's day-to-day work is an attractive way to assess performance. In-training evaluation (ITE) is “the process of observing and systematically documenting the on-going performance of a learner in real clinical settings during a period of training” [5]. This cost effective and subjective evaluation technique has the potential to formatively assess multiple competencies including history and physical exam skills, communication skills, team interaction, technical ability, and organizational domains. Although many areas may be assessed using in-training evaluation reports (ITER's), postgraduate surgery has traditionally relied upon them to assess technical skills using global rating scales [6]. At the end of training, accumulated ITER's are commonly summarized to construct the final ITER for summative assessment.

Despite their potential use, ITER's have been widely criticized particularly for their lack of reliability and validity [5, 7]. Validity may be compromised because traditional ITER's assess a restricted range of competencies and “halo effects”, where positive performance in one domain positively affects ratings in another, are common [5]. As well, when used without direct observation, the global rating scales introduce limited inter- and intra-rater reliabilities. Gray believes the difficulties have two causes [8]. One, faculty must be both teacher and assessor which are functions they receive little instruction in. Two, even with direct observation, documentation of target behavior is poor. ITERs are often neglected until the end of the rotation or months after. These two factors make ITERs and the final ITER both retrospective and subjective, thus limiting their utility in assessing technical skill.

Structured Methods of Assessment

With the stated limitations of current assessment methods in postgraduate surgical training, various alternatives have been investigated and developed. While numerous methods exist, the examples worthy of mention include objective structured clinical examinations (OSCEs), objective structured assessment of technical skills (OSATS), patient assessment and management examination (PAME) as well as emerging technologies including surgical simulation.

Objective Structured Clinical Examination

Pioneered by Harden et al. [9], the OSCE uses a series of stations revolving around a self contained clinical case.

Examinees rotate through each station, usually under a defined time limit, where they are presented with far ranging and standardized content.

Stations may be case based, involve standardized patients, require interpretation of diagnostic results, or involve minor technical procedures. Critical to this method is the use of trained assessors using standardized, validated, and objective criteria.

Unmistakable advantages of this method are the depth and breadth of material that may be examined, the ability to examine using simulated but realistic clinical environments, and improved reliability due to highly objective marking schemes. However, these advantages are tempered somewhat by both monetary and personnel expenses. While the estimated cost of using ITERs to evaluate 20 residents is \$10, OSCE estimates have ranged from several hundred to several thousand dollars *per candidate* [10]. As well, time restraints may preclude OSCE formats, as training expert assessors is a consuming process for busy clinicians that is associated with a real attrition rate.

Objective Structured Assessment of Technical Skills

Developed in Toronto by Reznick and colleagues, the Objective Structured Assessment of Technical Skills is a performance-based examination designed to assess the technical abilities of surgical learners [11]. OSATS requires the candidate to perform a range of surgical tasks while being graded against a validated checklist and global ratings sheet by multiple observers.

Eight stations involve bench model simulations of portions of general surgical procedures with a 15-min time limit [4]. Examiners assess against a global ratings scale with seven dimensions, each related to some component of operative performance (e.g., familiarity with operative procedure) and a “yes/no” checklist developed specifically for the task.

Although valuable in assessing technical ability in terms of knowledge and dexterity, the OSATS is currently limited in its ability to measure judgment due to the highly standardized tasks employed. Despite this and resource limitations similar to the OSCE, the OSATS is quite attractive and moving towards implementation in surgical training programs in some countries [3].

Patient Assessment and Management Examination

The Patient Assessment and Management Examination is a performance based clinical assessment that focuses on management, communication, and judgment abilities by utilizing eight clinical stations whose content is consistent with the American Board of Surgery (ABS) and the Royal College of

Physicians and Surgeons of Canada objectives for general surgery [12]. For each station, the candidate is given an introduction (referral letter and initial investigations), and they then interview, examine, and order investigations using a standardized patient. Results are subsequently given, and the standardized patient returns for counseling. Finally, the candidate is asked several predetermined questions relating to intraoperative events. Although its limitations mirror those of the OSCE and OSATS, data indicates that the combination of PAME and OSATS into a “comprehensive” examination is both valid and reliable for the formal assessment of readiness to practice [4].

Emerging Technologies: Assessment and Surgical Simulation

Although OSATS represents a considerable advance in the measure of technical skills, multiple pressures have stimulated the development of curricula to teach and evaluate fundamental skills in a laboratory setting. These include reduced resident work hours, increasing costs of operating room time, patient safety, the public and payers’ focus on medical errors, and the ethics of learning basic skills on patients. In response to these demands, simulators have been developed using inanimate box trainers and computer-based virtual reality platforms [13, 14]. The goals of these simulator-based curricula are to provide an opportunity to learn and assess basic skills in a relaxed and controlled environment so that a basic level of technical facility can be ensured prior to entering the operating room environment.

Advantages of training and assessment in skills labs include decreased stress compared to the operating room, the opportunity for repetitive practice and feedback, and the ability to tolerate, assess, and correct performance errors [15]. These advantages are particularly applicable to laparoscopy, an area where simulator curricula is attracting much interest because of the unique skills that must be learned by surgeons in training and surgeons in practice. This latter group has to develop a strategy to acquire novel skills and incorporate these skills into their clinical practice. Most of this retraining has been accomplished via mentorship on a limited number of real patients, and special training courses that require travel by either the mentor or retraining surgeon. A future alternative may incorporate surgical simulation using bench models, box trainers, and high fidelity simulators with mentorship on real patients after essential competencies are established in the lab. Unlike surgery with actual patients, practice in the simulator can be stopped at any time, allowing trainees an immediate chance to assess and correct mistakes while repeatedly performing challenging parts of procedures. It is also possible to improve the quality and detail of assessment feedback presented to the trainees by using video recordings of the performance, both during actual and simulated procedures.

It has been suggested that a lab based curriculum involving simulation for learning and assessment may lead to the effective development of expert performance [15]. With the early portion of a surgeon’s learning curve evaluated in a safe environment, a growing body of evidence now indicates that basic skills assessed against a criterion standard initially in the lab instead of on patients translates to superior performance with less errors in the OR environment [15–24]. Surely, appropriate assessment and feedback using computer technology and educational innovations such as simulation and bench model training are all factors influencing how future surgeons will practice the technical aspects of clinical skills [25]. At present, assessment formats using various technologies and simulation are being found both valid and reliable [26–30]. Although now controversial, these methods will likely 1 day be incorporated into current assessment models for credentialing surgeons.

Conclusion

Assessment is an integral component of training and credentialing surgeons for practice. Traditional methods of cognitive and technical appraisal are well established but have clear shortcomings. With knowledge of these deficits, surgeons, residency programs, and licensing bodies involved with evaluation may better able to refine existing methods of assessment and employ novel tools such as OSATS and PAME to ensure that their examination techniques are robust. Additionally, these stakeholders should attend to emerging technologies and simulation as they evolve thus ensuring a complete and sound assessment process in surgical education.

Compliance with Ethical Standards

Competing Interests The authors declare that they have no conflict of interest.

References

1. D’Costa (1986) The validity of credentialing examinations. *Eval Health Profess* 9:137–169
2. Brown (1992) Trends in assessment. In: Harden R, IR MH H (eds) *Approaches to the assessment of clinical competence part 1*. Center for Medical Education, Dundee
3. Darzi A, Mackay S (2001) Assessment of surgical competence. *Quality in Health Care* 10(Suppl 2):ii64–ii69
4. MacRae H, Regehr G, Leadbetter W, Reznick RK (2000) A comprehensive examination for senior surgical residents. *Am J Surg* 179(3):190–193
5. Turnbull J, Gray J, MacFadyen J (1998) Improving in-training evaluation programs. *Journal General Internal Medicine* 13(5): 317–323
6. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM (2004) Relationship between objective assessment of technical

- skills and subjective in-training evaluations in surgical residents. *J Am Coll Surg* 198(1):105–110
7. Wanzel KR, Ward M, Reznick RK (2002) Teaching the surgical craft: from selection to certification. *Curr Probl Surg* 39(6):573–659
 8. Gray J (1996) Global rating scales in residency education. *Acad Med* 71:S55–S63
 9. Harden R, Stevenson M, Downie W, Wilson G (1975) Assessment of clinical competence using objective structured examination. *Br Med J* 1:447–451
 10. Cusimano M, Cohen R, Tucker W, Murnaghan J, Kodama R, Reznick R (1994) Comparative analysis of the costs of administration of an OSCE (objective structured clinical examination). *Acad Med* 69:571–576
 11. Martin R, Reznick et al (1997) An objective structured assessment of technical skills (OSATS) for surgical residents. *The British Journal of Surgery* 84:273–278
 12. MacRae C, Regher et al (1997) A new assessment tool for senior surgical residents: the patient assessment and management examination. *Surgery* 122:335–343
 13. Feldman LSSV, Fried GM (2004) Using simulation to assess laparoscopic competence: ready for widespread use? *Surgery* 135:28–42
 14. Fried GM (2004) Simulators for laparoscopic surgery: a coming of age. *Asian J Surg* 27(1):1–3
 15. Ericsson KA (2004) Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 79(10 Suppl):S70–S81
 16. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M et al (1999) Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg* 177(2): 167–170
 17. Grober EDHS, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA (2004) The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Ann Surg* 240:374–371
 18. Grober EDHS, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS, Jarvi KA (2004) Laboratory based training in urological microsurgery with bench model simulators: a randomized controlled trial evaluating the durability of technical skill. *J Urol* 172:378–381
 19. Naik VN, Matsumoto ED, Houston PL, Hamstra SJ, Yeung RY, Mallon JS et al (2001) Fiberoptic orotracheal intubation on anesthetized patients: do manipulation skills learned on a simple model transfer into the operating room? *Anesthesiology* 95(2): 343–348
 20. Grantcharov TP, Schulze S, Kristiansen VB (2007) The impact of objective assessment and constructive feedback on improvement of laparoscopic performance in the operating room. *Surg Endosc* 21(12):2240–2243
 21. Seymour NEGA, Roman SA (2002) Virtual reality training improves operating room performance. *Ann Surg* 236:458–464
 22. Taffinder NSC, Fishwick RJ, McManus IC, Darzi A (1998) Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. *Stud Health Technol Inform* 50:124–130
 23. Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G et al (2004) Proving the value of simulation in laparoscopic surgery. *Ann Surg* 240(3):518–525 discussion 525–8
 24. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C et al (2009) Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *BMJ* 338:b1802
 25. Reznick RK, MacRae H (2006) Teaching surgical skills—changes in the wind. *N Engl J Med* 355(25):2664–2669
 26. Kundhal PS, Grantcharov TP (2009) Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. *Surg Endosc* 23(3):645–649
 27. Larsen CR, Grantcharov T, Aggarwal R, Tully A, Sorensen JL, Dalsgaard T et al (2006) Objective assessment of gynecologic laparoscopic skills using the LapSimGyn virtual reality simulator. *Surg Endosc* 20(9):1460–1466
 28. Grantcharov TP, Carstensen L, Schulze S (2005) Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. *Jsls* 9(2):130–133
 29. Creco E, Grantcharov, Okrainec (2008) Do performance measures on the LapMentor VR simulator predict FLS performance? *Can J Surg* 51(Suppl):14
 30. McCluney AL, Vassiliou MC, Kaneva PA, Cao J, Stanbridge DD, Feldman LS et al (2007) FLS simulator performance predicts intraoperative laparoscopic skill. *Surg Endosc* 21(11):1991–1995