Thomas Jefferson University
Jefferson Digital Commons

Department of Obstetrics and Gynecology Faculty Papers

Department of Obstetrics and Gynecology

June 2006

The objective structured assessment of technical skills and the ACGME competencies

Carmen J. Sultana
Thomas Jefferson University, Carmen.Sultana@jefferson.edu

Let us know how access to this document benefits you

Follow this and additional works at: http://jdc.jefferson.edu/obgynfp

Part of the Obstetrics and Gynecology Commons

Recommended Citation
Sultana, Carmen J., "The objective structured assessment of technical skills and the ACGME competencies" (2006). *Department of Obstetrics and Gynecology Faculty Papers.* Paper 1.
http://jdc.jefferson.edu/obgynfp/1

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's Academic & Instructional Support & Resources Department (AISR). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Obstetrics and Gynecology Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

The Objective Structured Assessment of Technical Skills and the ACGME Competencies

Carmen J. Sultana, M.D.
Department of Obstetrics and Gynecology
Jefferson Medical College
Philadelphia, Pennsylvania

Corresponding author:

Carmen J. Sultana MD
Department of Obstetrics and Gynecology
834 Chestnut Street, Suite 400
Philadelphia PA 19107
215-955-9217
Fax 215 955-5041
carmen.sultana@jefferson.edu

Keywords: objective, structured assessment of technical skills; urogynecology; surgical training; residency education; surgery; bench models

Synopsis

Objective structured assessments of technical skills (OSATS) are structured operating room or laboratory assessment of residents' surgical skills. They can be used to evaluate and teach both basic and complex skills to residents. The literature on their use is reviewed. Future use of virtual reality simulators is discussed.

The Objective Structured Assessment of Technical Skills and the ACGME Competencies

Assessment drives both how a subject is taught and what is taught. The Accreditation Council for Graduate Medical Education (ACGME) Outcomes Project has provided a new impetus for residency training programs to examine and improve the way in which they teach technical skills to residents and assess the acquisition of those skills. Pressure from hospitals and other organizations on residency programs to credential residents in specific procedures has emerged as another factor driving change in both surgical teaching and evaluation. The ACGME mandate to assess competency –in this case, surgical skills falls under patient care and medical knowledge-means that programs will be looking for tools with which to measure the technical skills of their residents. The OSAT (Objective Structured Assessment of Technical Skills) is a variation on the OSCE, or Objective Structured Clinical Examination. It is one of the newer tools used to assess competency. The OSAT is cited in the ACGME Evaluation Toolbox on the website as the most desirable evaluation tool for the Patient Care topics including interviewing, counseling, preventive services, and performance of physical examinations. Simulators and models are listed as most desirable tools for evaluating medical procedures. An OSAT combines both. This chapter will review the history of teaching technical skills, components of surgical skills assessment, definition of an OSAT, and suggestions for future development of surgical skills assessment tools.

Why evaluate or teach outside the operating room setting? Until recently, the cognitive apprenticeship model of learning , in which residents learned by working side by side with experienced surgeons in a real life setting, was the major method with which to impart surgical skills. The framework for this transformation of novice to expert exploits several components. It relies on role modeling and relationships with the instructor who coaches the learner. A scaffolding or framework of basic skills is built upon as learners articulate or string together the steps of the procedures being taught, eventually acquiring the ability to perform progressively more complex procedures. The apprentice model is limited by the availability of mentors and their teaching ability, the variable types of procedures and patients that are available for training, and the important issues of how to maintain patient safety and the quality of care while attending to the needs of the learner.

Dr. R. Reznick, whose group has many publications on the subject of teaching and evaluating surgical skills, noted that the "Most important ingredient (of surgical teaching) is the appreciation of the importance of … skill acquisition accompanied by the access to a knowledgeable patient faculty" [1]. He cites the importance of perception, integration and automatization [2]. Perception is the awareness of the essentials of procedures; integration refers to learning of the framework and steps; and automatization occurs after repetition. In addition to these concepts, those teaching residents should also be aware of the ways in which adults learn. Adults set goals and objectives for skills they wish to acquire which affects their interest in a given subject. They tend to focus on tasks rather than on abstract knowledge. They also display metacognition, or an awareness of

their own learning strategies and preferences. In keeping with this, the instructor of adult learners needs to be centered on the learner rather than solely on the subject material. He or she should assess the learner's level of skill brought to the situation so that the new knowledge can be integrated with knowledge or skills previously learned. Transferable concepts should be emphasized. Formative evaluation, or ongoing feedback to the learner, is also important.

In creating a new approach to surgical assessment, three components should be addressed: feasibility, reliability and validity. Evaluation methods that are feasible can be done easily and inexpensively. Reliable methods are reproducible and precise. Validity can be thought of as "are we measuring what we intend to?" A valid test is predictive, or forecasts future performance and has appropriate content, which, for the assessment of surgical skills, means that it would measure technical" skill" rather than medical "knowledge". In addition, a valid test is concurrent if it correlates with some gold standard, and has construct validity or the ability to separate groups. Finally, a valid test must have face validity or resemble the real world.

How can we incorporate this information into the development of an assessment method? Commonly used global rating forms do not accurately discriminate skills and are ranked low in preferred methods by the ACGME. Watts et al [3] evaluated procedure logs, direct observation, direct observation with criteria, animal models with criteria, and videotapes in terms of reliability and validity. They gave the highest rating to direct

observation with criteria, noting that the performance of models and videotaping depended on the degree of realism achieved.

One method of direct observation with criteria is the Global Rating Scale of Operative Performance that is part of the ACGME Toolbox [4] (Figure 1). The components assessed include respect for tissues, time and motion, instrument handling, knowledge of instruments, flow of the operation, uses of assistants, and knowledge of specific procedure on a five point Likert scale. The same author went on to develop a structured technical skills assessment form (STSAF) that partitioned particular procedures into their smallest fundamental components [5]. This was done for cholecystectomy, hernia repair and bowel repair. It included a ten point summary scales, had high levels of inter-rater reliability and was able to differentiate junior from senior residents, suggesting construct validity.

The logical next step for investigators was to control the assessment environment by doing structured evaluations in a bench setting with models instead of in the operating room. An initial trial in 1996 used eight 15 minute bench stations for simulated skin excision, T-tube insertion, abdominal wall closure, bowel anastomosis, IVC hemorrhage, pyloroplasty, and tracheostomy [4]. The exercise required 48 surgeon examiners to spend 2 hours each to examine all their residents, with a cost of abut $200 Canadian per resident. They established an interstation reliability of 0.843 for the global rating and

0.781 for the checklist. ANOVA showed a significant difference according to training level, implying construct validity.

The same group of investigators then set out to compare the use of live animals to the bench models [6]. Again, they used 48 examiners who spent three hours each to examine 20 residents, using six stations. All residents went through both types of stations. Three ratings methods were used: the checklists, the global rating form, and pass/fail judgments. The ANOVA results showed that the scores for the two examiners correlated for all three methods. The reliability indices for the checklist and global scores were moderate to high. Internal consistency was moderate to high for the live checklist. Both formats were equivalent for the global rating. The global rating was better at differentiating resident levels. They noted that they would need more stations to have higher reliability. The conclusion was that the bench models were equivalent to the use of anesthetized animals, at a cost of $160 versus $600. In another study, they compared bench models to cadavers and textbook training in six procedures using the same checklists and global rating scales and found that cadaver and bench models were superior to textbook teaching. [7]

Other investigators have similar findings in an OB/GYN program. Goff et al [8] described pig lab teaching sessions on knot tying, suturing, tying on a passer, laparoscopic tasks using pegs, and transferring objects from one instrument to another.

The animal labs included tubal ligation, ostomy, salpingectomy, open repair of cystotomy, enterotomy, salpingo-oophorectomy, hypogastric artery ligation and more advanced procedures. A written pre-test and post-test six months later was used. The subjective faculty evaluation differentiated resident training levels and residents felt the labs improved their skills. Subsequent studies included a seven station, four hour, $1500 pig laboratory. The procedure included laparoscopic port placement, salpingostomy, suturing, vessel ligation and open hypogastric artery ligation, enterotomy repair and salpingo-oophorectomy. Residents were evaluated with a task specific checklist, global rating scale and pass-fail grade. They also used self scored and skills ratings by the faculty. [9] The reliability and construct validity of the evaluations were high. The next two studies by this group used bench models instead of live animals. Lentz et al [10] used a written test and 12 stations including six laparoscopic procedures, four abdominal, and a knot tying station at a cost of $50 each. The residents were scored on time and a five point scale by two examiners. The results showed that the global ratings performed better than the checklists. Reliability varied with the task, but overall was 0.79. The construct validity was best with the total score. This group later repeated this study in a blinded fashion with seven stations, and found similar results for construct validity and reliability [11]. One examiner of an examiner pair for each resident did not know the resident. The feasibility of this evaluation instrument was further examined by administering a version of their OSAT examination to a total of 116 residents from five different programs [12]. One of the programs was the site where they had carried out the original exercises in teaching and evaluation. The combination of one blinded and one unblinded examiner was again used, with an overall inter-rater reliability of .95 (range, .71 to .97). The test

consisted of three open and three laparoscopic tasks, and took each resident 90 minutes. Both the global rating and checklists discriminated between resident levels. The equipment costs were between $40-150 per resident depending on the tasks chosen.

Finally, residents who had experienced the groups' laboratory based curriculum performed better in terms of scores and time to complete tasks than those who were naïve to the laboratory setup.

Other trials are few; Coleman et al conducted a randomized trial of videotaped laparoscopic salpingectomy after intensive video skills training and noted improvement in timed drill scores in the drill versus control group [13]. Reliability and validity were established for an OSAT for episiotomy repair that evaluated six components of the repair and seven global surgical skills as well as a pass/fail assessment [14].

In addition to global evaluations and checklists, dexterity analysis systems may in the future provide another level of evaluation [15]. One example, the Isotrack (Polhemus, United States) is an electromagnetic field generator with sensors that attach to the surgeon's hands and collect information on time and hand movements in performing tasks. Construct validity of this system has been demonstrated, and it has been used in assessment of laparoscopic and open tasks [16, 17]. Virtual reality systems such as the MIST-VR minimally invasive surgical trainer was developed through a task analysis of a surgical procedure, then replicated in the virtual domain. Metrics such as economy of

movement, length of path, and instrument errors have been used to validate the assessment of basic laparoscopic skills [18, 19].

In summary, bench models for testing surgical skills appear to be effective tools for identifying and differentiating different levels of resident skills at an inexpensive cost. Many examples of specific models are in the literature or presented at meetings. The evaluation can be incorporated into a surgical curriculum that programs can tailor to their own needs. Resources include the ACGME Outcomes Project link at www.acgme.org as well as the CREOG Educational Objectives and Surgical Curriculum, via www.acog.org .

1. Reznick RK. Teaching and Testing Technical Skills. Am J Surg 1993; 165: 358-361.
2. Kopta JA. An approach to the evaluation of operative skills. Surgery 1971;70:297-303.
3. Watts J, Feldman WB. Assessment of technical skills. In: Neufeld VR, Norman GR, ed. Assessing clinical competence. New York: Springer, 1985: 259-74.
4. Reznick RK, Regehr G, MacRae H et al. Testing Technical Skill via an Innovative “Bench Station“ Examination. Am J Surg 1996;172:226-30.
5. Winckel CP, Reznick RK, Cohen R et al. Reliability and Construct Validity of a Structured Technical Skills Assessment Form. Am J Surg 1994;167:423-427.

6. Martin JA, Regehr G, Reznick RK et al. Objective Structured Assessment of Technical Skill (OSATS) for Surgical Residents. Br J Surg 1997;84:273-78.
7. Anastakis D, 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
8. Goff BA, Lentz GM; Lee DM, Mandel LS. Formal Teaching of Surgical Skills in an Obstetric-Gynecologic Residency. Obstet Gynecol 1999;93:785-90.
9. Goff BA, Lentz GM; Lee DM et al. Development of an objective structured assessment of technical skills for obstetric and gynecology residents. Obstet Gynecol 2000;96: 146-50.
10. Lentz GM, Mandel LS; Lee DM et al. Testing surgical skills of obstetric and gynecologic residents in a bench laboratory setting:Validity and reliability. Am J Obstet Gynecol 2001;184:1462-1470.
11. Goff BA, Nielson PE, Lentz GM. et al. Surgical skills assessment: a blinded examination of obstetrics and gynecology residents. Am J Obstet Gynecol 2002;186:613-7.
12. Goff B, Mandel L, Lentz G et al. Assessment of resident surgical skills: Is testing feasible? Am J Obstet Gynecol 2005;192:1331-40.
13. Coleman RL, Muller CY. Effects of a laboratory-based skills curriculum on laparoscopic proficiency: a randomized trial. Am J Obstet Gynecol 2002;186:836-42.
14. Nielsen PE, Foglia LM, Mandel LS, Chow, GE. Objective structured assessment of technical skills for episiotomy repair. Am J Obstets Gynecol 2003;1891257-60.

15. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ 2003, 327:1032-7.
16. Datta V, Mackay S, Mandalia M. Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Amer Coll Surg 2001;193:479-85.
17. Mackay S. et al. Multiple Objective Measures of Skill (MOMS) a new approach to the assessment of technical ability in surgical trainees Annals Surg 2003;238:291-300.
18. Taffinder N, Sutton C, Fishwick RJ et al. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomized controlled studies using the MIST-VR laparoscopic simulator. Stud Health Technol Inform 1998;50:124-30
19. Bann S, Kwok KF, Lo CY, et al. Objective assessment of technical skills of surgical trainees in Hong Kong. Br J Surg 2003;90:1294-1299.