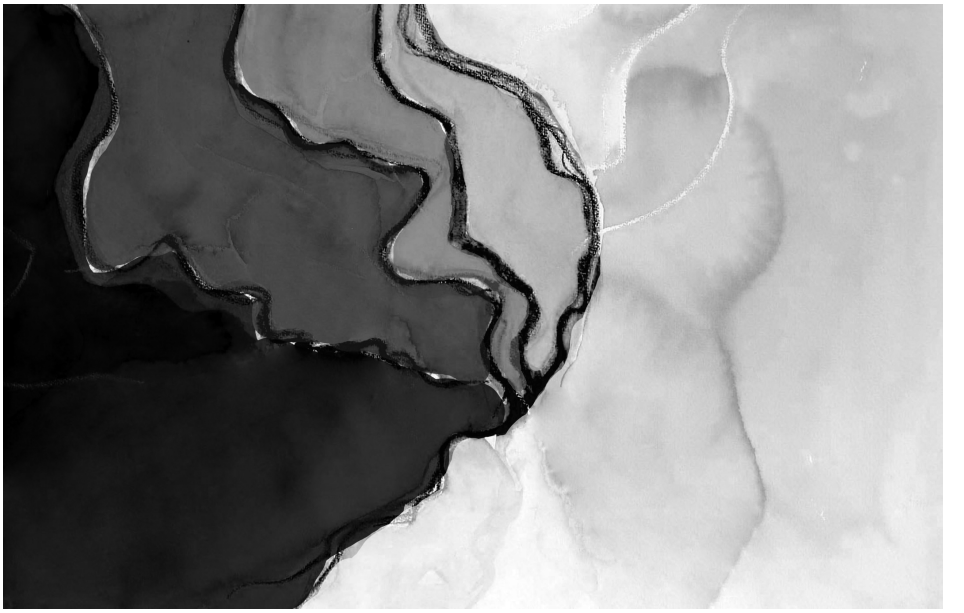
Acquiring minimally invasive surgical skills

Hiemstra, E.

CHAPTER 9
IMPLEMENTATION OF OSATS IN THE RESIDENCY PROGRAM: A BENCHMARK STUDY



Ellen Hiemstra
Navid Hossein pour Khaledian
J. Baptist M.Z. Trimbos
Frank Willem Jansen

Submitted

INTRODUCTION

The exposure to surgical procedures during residency training has decreased. Consequently, residents have to achieve the same competencies in fewer working hours than their counterparts some decades ago.[Haluck & Krummel, 2000] The decrease in working hours is exaggerated by a trend to non-surgical therapies for certain traits.[Hammond & Karthigasu, 2006] With this development, the apprenticeship model relying on experiential training, large number of procedures performed and subjective, observational assessment of surgical skills no longer suffices. Instead, a different and highly efficient program is required to achieve surgical proficiency.

However, a large survey in 1998 in the United States revealed that evaluation of surgical skills is usually done by subjective faculty assessment at the end of a rotation, and, therefore, is based on recollection of events over time.[Mandel et al., 2000] Obviously, this method lacks reliability and validity. Therefore, there is a pressing need to evaluate surgical skills more objectively in a structured fashion.

In response, assessment tools have been developed to evaluate surgical skills.[Reznick et al., 1997] An example of such a tool is the OSATS (Objective Structured Assessment of Technical Skills). OSATS assesses discrete domains of surgical competence. It has proven to be reliable and construct-valid in bench models and in live animal models.[Martin et al., 1997] Goff et al. showed the reliability and validity of OSATS in animal models for residents in Obstetrics and Gynaecology.[Goff et al., 2000]

Training on bench models in skills laboratories enables repeated skills training without the risk of harming patients. However, practising on a bench model is not equivalent to performing surgery on a living patient in the operating room (OR). Obviously, a surgeon has to become proficient in the latter. As a consequence, scientific evidence of more objective intraoperative assessment tools is needed. In fact, the construct validity of intraoperative use of OSATS has been proven in two studies[Aggarwal et al., 2008; Hiemstra et al., 2011]. Additionally, residents and supervisors perceive intraoperative administration of OSATS to be a valuable and valid tool, as was revealed from a questionnaire in United Kingdom.[Bodle et al., 2008]

However, prior to large-scale implementation of assessment with OSATS, more information is required.[Bodle et al., 2008] It has been stated that an objective assessment tool can be used for authorization,[Darzi et al., 1999] but cut-off values have not been defined.[van Hove et al., 2010] Additionally, it is unclear whether residents' self-assessment is in accordance with their supervisor's rating using the OSATS. This can be interpreted as a form of inter-rater reliability, and is important in the renewal of residency programs with an increasing focus on self-assessment. Finally, we are interested in the opinion of clinicians who will have to work with intraoperative administration of the OSATS after its implementation.

The aim of our prospectively designed study is threefold; first, to establish at which OSATS score a supervisor judges a resident able to perform the procedure autonomously; second, to evaluate the reliability between resident and supervising gynaecologist regarding intraoperative assessment of technical surgical skills, and finally, to question aspects of the satisfaction of residents and supervisors with the intraoperative administration of OSATS.

MATERIALS AND METHODS

To answer the research questions, study-specific OSATS forms were distributed for application in clinical practice. Additionally, a survey was performed among users of the OSATS.

Intraoperative administration of OSATS

All obstetrics and gynaecology (Ob/Gyn) residents who were attending their specialist training at the Leiden University Medical Center (LUMC) or at their affiliated teaching hospitals from July 2007 to June 2009, were asked to participate in the study. They were informed by means of mailing and by an individual briefing. Residents were asked to complete a study-specific double-sided assessment form after each procedure they had performed as the primary surgeon. One side had to be filled out by the resident, and the other side by the supervising consultant. Each side of the form contained a general global rating scale of the OSATS and the question whether the resident would have been able to perform the procedure autonomously, i.e. without supervision (Figure 1). On the global rating scale, which had been adapted from Martin et al.[Martin et al., 1997], six domains of technical surgical skills could be rated on a 1 to 5 scale in which ratings 1, 3 and 5 had an explicit description. The ability to perform a certain procedure autonomously was rated on a 3-points scale (no, maybe or yes).

After the self-assessment on one side of the form, the supervisor had to complete the other side blind for the results of the self-assessment. Next, the results were discussed with

OSATS - global rating scale of operative performance

Please circle the number corresponding to the candidate's performance in each category, irrespective of training level.

Respect for Tissue:	1 Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments	2	3 Careful handling of tissue but occasionally caused inadvertent damage	4	5 Consistently handled tissues appropriately with minimal damage
Time and Motion:	1 Many unnecessary moves	2	3 Efficient time/motion but some unnecessary moves	4	5 Clear economy of movement and maximum efficiency
Knowledge and handling of instrument:	1 Lack of Knowledge of Instruments	2	3 Competent use of instruments but occasionally appeared stiff or awkward	4	5 Obvious familiarity with instruments
Flow of operation:	1 Frequently stopped procedure and seemed unsure of next move	2	3 Demonstrated some forward planning with reasonable progression of procedure	4	5 Obviously planned course of procedure with effortless flow from one movement to the next
Use of assistants:	1 Consistently placed assistants poorly or failed to use assistants	2	3 Appropriate use of assistants most of the time	4	5 Strategically used assistants to the best advantage at all times
Knowledge of specific procedure:	1 Deficient knowledge. Needed specific instructions at most steps	2	3 Knew all important steps of procedure	4	5 Demonstrated familiarity with all aspect of operation
Is the resident able to perform the procedure autonomously?			No Maybe Yes Please circle the most appropriate answer		

Figure 1. OSATS form used for the study.

the resident in order to create a learning opportunity. The supervisor could be any consultant gynaecologist working as a staff member at one of the hospitals of the study. They had uniformly been instructed to fill out the OSATS form by rating the resident as objectively as possible, irrespective of his/her level of experience.

Questionnaire

Additionally, a survey was held among the users of the OSATS in July 2009, regarding the user's satisfaction with the intraoperative assessment tool. This questionnaire was sent to residents who were attending their Ob/Gyn residency training at that time and to the consultant gynaecologists who actively participated in teaching surgical skills in the ORs of these hospitals. A Likert scale was used to have the respondents express their agreement or disagreement with five statements on OSATS on a five-point scale. The five statements included that OSATS is a valid instrument, it is subjective (or objective), it should be used for assessment, it helps in acquiring surgical skills, and it leads to irrelevant paperwork. Finally, the participants were asked for their opinion on the ideal frequency of administration of an OSATS.

Statistics

The results were collected in the statistical SPSS program (SPSS, version 16, SPSS inc., Chicago, IL). The corresponding median OSATS scores were calculated for a positive, uncertain and negative response to the question of whether a resident was able to perform the procedure autonomously. The intraclass correlation coefficient (ICC) for single measurements was used to determine the inter-observer reliability. Although arbitrary, a frequently used nomenclature for the ICC is that a score ranging from 0.41 to 0.60 indicates moderate agreement, a score ranging from 0.61 to 0.80 substantial agreement, and a score above 0.81 indicates perfect agreement. [Landis et al., 1977] However, the use of ICC can be deceptive, as the outcome is also dependant of the number of items scored. However, we have chosen this measure to enable a comparison with results of other studies about the reliability of the OSATS.

RESULTS

The participants were 19 residents, equally distributed among all six postgraduate years. Six were male, thirteen were female. In total, 127 study forms were collected and the data analysed. The procedures assessed related to abdominal, vaginal, laparoscopic and hysteroscopic surgery. For an individual resident, the median number of procedures assessed was 14 (range 1-28). A total of 27 gynaecologists assessed these procedures.

The OSATS score had been completed by the gynaecologist in all the forms returned (100%). In 122 cases (96%), the question about the ability to perform the procedure autonomously was also completed. The OSATS score on the resident's site was filled out in 123 cases (97%), and the ability question was filled out in only 92 cases (75%).

The box plots of the OSATS scores corresponding with the ability to operate autonomously are presented in figure 2. The median OSATS score that corresponded with the supervisor's opinion that the resident was able to perform the procedure autonomously was 28 (range 20-30). This corresponds with 92% of the maximum score, taking the possible range from 6 to 30 into account.

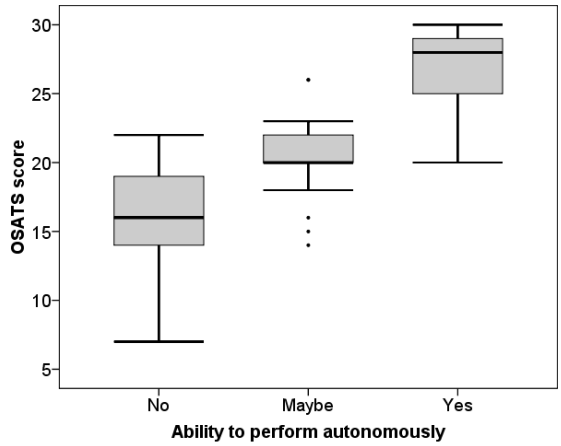


Figure 2. OSATS score against the ability to perform autonomously.

Box plots of the OSATS scores as rated by supervisors against their answer on the question whether to resident is able to perform the procedure autonomously. The lower and upper lining of the boxes represent the 25th and 75th percentile, the black line is the box is the median. The number of procedures rated with no, yes and maybe were 58, 36 and 28 resp.

The ICC of the total OSATS score of residents and supervisors was 0.78 (95% CI 0.70-0.84), indicating substantial agreement. Additional analysis was carried out on the cases in which residents overrated versus underrated their performance (defined as a resident's OSATS score of ≥ 3 versus ≤ 3 points compared to the supervisor's rating). According to this definition, overrating was present in 15% (n=18), agreement in 53% (n=66), and underrating in 32% (n=39) of the procedures (Figure 3). In seven of 18 cases (38%) in which the resident overrated his/her performance, the resident was in the last two years of residency training (PGY5 or PGY6). Next, supervisors and residents agreed in 64 out of 91 (70%) cases regarding the question of whether

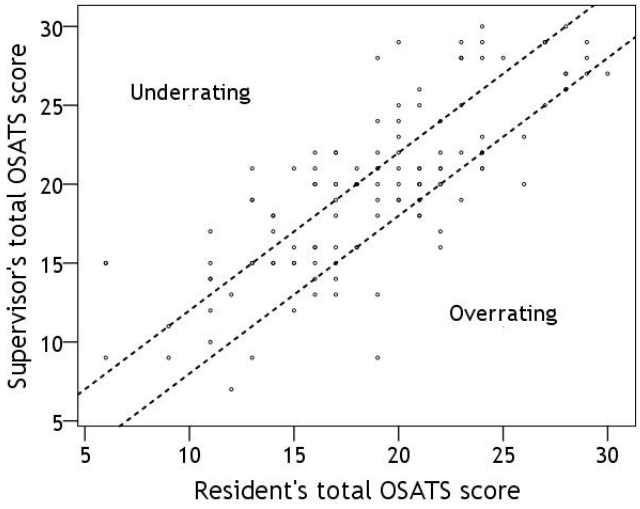


Figure 3. Supervisor's OSATS score plotted against Resident's OSATS score. Scatter plot. The area between the dotted lines represents agreement (a difference of 2 or less points). Under- and overrating of the resident is defined as a self-perceived OSATS score of at least 3 points lower resp. higher than the supervisor's score.

a resident was able to perform the procedure autonomously. No absolute disagreement occurred (Table 1).

Table 1. Agreement resident and supervisor.

Agreement Supervisor and Resident regarding ability to perform procedure autonomously					
		Ability according to resident			
Count		No	Maybe	Yes	Total
Ability according to supervisor	No	38	12	0	50
	Maybe	5	9	5	19
	Yes	0	5	17	22
	Total	43	26	22	91

The response rates to the questionnaire were 96% and 100% respectively: 23 residents and 15 supervisors returned the questionnaire. One resident did not respond due to inexperience with the OSATS because of having just started the residency training. The results are presented in table 2. Regarding the ideal frequency of administration of an OSATS, the majority of the residents (17 out of 23) answered that this was after each procedure they had performed as the primary surgeon.

Table 2. Results survey.

OSATS ...		Median score residents	Median score supervisors
is a valid instrument	(1= strongly disagree; 5=strongly agree)	3	3
is (subjective/objective)	(1=subjective; 5=objective)	2-3	3
should be used for assessment	(1=strongly disagree; 5=strongly agree)	4	4
helps acquiring surgical skills	(1=strongly disagree; 5=strongly agree)	3-4	3-4
leads to irrelevant paperwork	(1=strongly disagree; 5=strongly agree)	2-3	2

DISCUSSION

At 92% of the maximum OSATS score (28/30 points), a supervisor found a resident Ob/Gyn is able to perform a procedure autonomously, irrespective the kind of procedure. This is a rather high score, especially in comparison with a randomized clinical trial (RCT) that described a cut-off point for a certain level of proficiency.[Bijen et al., 2009] The authors used 75% of the maximum OSATS score (28/35 points) to select surgeons as proficient to perform a total laparoscopic hysterectomy in the case of endometrial cancer. The high score measured in this study might be due to the supervisor's striving for perfection and his/her hesitancy to authorize

the resident to perform the surgery autonomously. Furthermore, it may also be the result of the desire to add something to a resident's learning process, and not wanting to be redundant as a teacher. Also, an assessing supervisor in our study may have interpreted the maximum score as the level required at the end of residency, while an assessor during the clinical trial[Bijen et al., 2009] may have taken the maximum score as absolute perfection. Assessing an individual with a certain frame of reference in mind diminishes the objectivity the assessment instrument.

Substantial agreement is present between residents' and supervisors' OSATS score. This is in accordance with the inter-observer reliability that is found in laboratory settings,[Martin et al., 1997] and endorses the quality of OSATS as intraoperative assessment tool. Residents, however, cannot be regarded as equivalent to observers since they performed a self-assessment. Regarding self-assessment, the study of Mandel et al. found good reliability comparing the rating of trained faculty observers with residents' self-assessment of surgical skills on bench models using the OSATS.[Mandel et al., 2005] However, a recent review raises questions about the abilities of health professionals to generate accurate judgments of their own performance.[Eva et al., 2005] Even in concrete areas such as technical knowledge and ability, the self-assessment was found to be inaccurate.[Ginsburg et al., 2000; Gordon, 1992; Hodges et al., 2001] An issue of greater concern is that those who perform worst on external assessment may also overrate their performance on self-assessment.[Davis et al., 2006; Hodges et al., 2001; Kruger et al., 2006; Lynn et al., 2006] Fortunately, we showed that only a minority (15%) overrated themselves. Nonetheless, overconfidence is dangerous, especially when combined with suboptimal performance at the end of residency training. Notwithstanding, self-assessment has assumed increasing importance, though external assessment will always play an essential role during the process of acquiring certain skills.[Mandel et al., 2005]

Official assessments generate paperwork. However, neither residents nor supervisors agreed with the statement that OSATS leads to irrelevant paperwork. Instead, they agreed that it should be used for assessment and evaluation, and that it helps to improve residents' surgical skills. Furthermore, residents state that they want to be assessed after every surgical procedure they performed as primary surgeon. In daily practice, however, they only request this during a minority of the procedures performed. Probably, their answers express socially desirable behaviour, or there may be barriers present that discourage residents from asking for an assessment with an OSATS. A possible explanation is that practical impediments (e.g. insufficient time) hamper the frequency of administration of the OSATS. On the other hand, it is questionable whether all procedures need to be assessed. In our opinion, regular assessments distributed over time are sufficiently able to show a resident's skills level as well as the expected progression in skills and in the performance of the surgery. Worthwhile situation-specific feedback and advice need to be given during every surgical procedure. Strikingly, none of the participants judged the OSATS to be a very objective assessment tool. This is in contradiction with the objectivity the OSATS stand for. Obviously, no judgment from one person about another will be free from subjectivity. Though, in our opinion, this finding suggests that someone's surgical performance cannot automatically be derived from an OSATS score.

This study addresses a very difficult area of surgical skills evaluation, and was conducted under regular conditions mirroring daily practice in a residency program. Following such a

design, real-life influences were allowed to colour the study results. Individual variation will have been present among residents, supervisors and procedures. It is unlikely that all residents had the same level of motivation to participate in the study. This was illustrated by the large range of number of OSATS each resident had collected. Additionally, variation will have been present in the extent to which supervisors allowed residents to independently perform a surgical procedure and in their method of assessment. Also, the procedures assessed will have varied widely with respect to difficulty and the risk of complication. However, the OSATS will be implemented in the actual clinical situation and not under predefined study conditions. This expresses the strength and the value of this study.

Training in all specialties is evolving and moving towards more competency-based outcome measurements rather than solely based on the length of training. This is a positive development. However, we should avoid an indiscriminate implementation of instruments such as the OSATS, especially, with respect to drawing consequences to certain scores like authorization. Authorization is a more complex, multifactor process. During this process, the importance of an OSATS is limited. Other competencies also have to be taken into account, such as knowing when to operate and when not, recognizing someone's own abilities and inabilities, asking for help when needed, and being open to suggestions from colleagues. Therefore, all these competencies should be evaluated prior to authorization. Finally, acquiring surgical proficiency is an ever-continuing process that does not end with the completion of residency training.