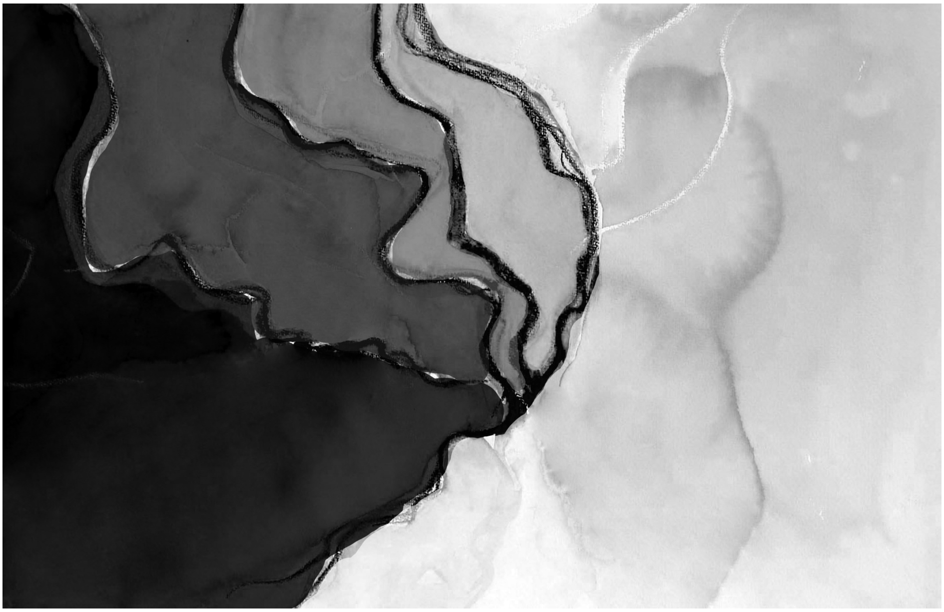
Acquiring minimally invasive surgical skills

Hiemstra, E.

CHAPTER 8
THE VALUE OF AN OBJECTIVE ASSESSMENT TOOL IN THE OPERATING ROOM



Ellen Hiemstra
Wendela Kolkman
Ron Wolterbeek
J. Baptist M.Z. Trimbos
Frank Willem Jansen

Adapted from Can J Surg, 2011; 54: 116-22

INTRODUCTION

Nowadays, it is becoming more and more difficult to achieve surgical proficiency. Residents experience less training due to reduced working hours and a decreased surgical caseload.[Hammond et al., 2006] Additionally, with the development of new surgical techniques, skills acquisition is more challenging.[Haluck & Krummel, 2000] Currently, basic surgical procedures are sufficiently mastered after finishing residency training, but advanced procedures are not.[Kolkman et al., 2006] Ultimately, skills deficiencies will impede post-residency performance.[Shay et al., 2002] Moreover, residency programs still rely heavily on informal and subjective evaluations based on recollections of supervisors.[Kolkman et al., 2005; Mandel et al., 2000] Therefore, on one hand, surgical skills training need to become more efficient, and on the other hand, appropriate assessment is required in order to optimally benefit from the spare learning moments in the operating room (OR).

An objective assessment tool can fulfil an important role during operative training.[Aggarwal et al., 2008; Beard, 2007] Such a tool can be an aid to the learning process through constructive feedback on performance. Secondly, an assessment tool can be applied to establish competency levels and to mark progression in time. Finally, it can provide a benchmark criteria to be used as a training goal or for credentialing purposes.[Cuschieri et al., 2001; Darzi et al., 1999]

To fulfil this need for an objective assessment tool, the OSATS (Objective Structured Assessment of Technical Skills) was developed by Martin et al. in Toronto in 1997.[Martin et al., 1997] An OSATS consists of a procedure-specific checklist, a pass/fail judgment and a global rating scale. The latter turned out to be superior in terms of reliability and validity.[Goff et al., 2000; Martin et al., 1997; Swift et al., 2006] On this global rating scale, domains are scored on a 1 to 5 Likert-scale, with an explicit description at point 1,3 and 5.

So far, studies about the quality of OSATS have mainly been conducted in simulators or live animal models.[Reznick et al., 2006] Although applying OSATS in simulator settings has the benefit that repeated practice is enabled without the risk to harm patients, simulators will never perfectly mimic operative conditions. Therefore, OSATS have been implemented for the assessment of real surgical procedures on a large scale in the Netherlands in residency programs. Moreover, plans are being developed to use this form of assessment tool for certification purposes after residency training. However, only a few studies have investigated the value of intraoperative use of OSATS.[Aggarwal et al., 2008; Bodle et al., 2008] Aggarwal et al. found that the OSATS score discriminates between a novice and an expert surgeon performing a laparoscopic cholecystectomy demonstrated by video-based assessment.[Aggarwal et al., 2008] Bodle et al. concluded from feedback questionnaires that trainers and trainees in the United Kingdom perceived the OSATS to be valid and valuable.[Bodle et al., 2008] In the absence of data on the implementation of OSATS in daily practice, the current study was conducted in order to assess its value in clinical practice by analysing residents' learning curves for a variety of surgical procedures in gynaecology.

MATERIALS AND METHODS

In the Netherlands, the Obstetrics and Gynaecology (Ob/Gyn) residency program lasts six years. On average, three of these six years are spent in a university teaching hospital, and the complementary period is spent in a non-university teaching hospital. The university hospitals provide a curriculum to train residents in a variety of subspecialties, like reproductive health care, perinatology and oncology. Specifically, a three-month clinical rotation is spent on gynaecological surgery. During this rotation, which is generally attended during the fourth postgraduate year (PGY), residents are scheduled to perform surgery in the OR for four days a week. Gradually, a resident is given more responsibility as experience accrues, depending on the resident's technical skills, the type of procedure and patient characteristics. Finally, a resident performs a procedure as the primary surgeon, in the presence of a supervising consultant.

Study Design

In 2005, the global rating scale of the OSATS (referred to as "OSATS" in this thesis) was introduced at the department of Ob/Gyn of the Leiden University Medical Center in an observational study of its implementation in clinical practice (Figure 1). The assessment tool had been adapted from Martin et al.[Martin et al., 1997] The six domains of an OSATS represent aspects of technical competence in surgery. The only modification to the original form is that we merged the domains 'knowledge of instruments' and 'instrument handling'. This is in accordance with the version of the OSATS form used by the Royal College of Obstetrics and Gynaecology.[RCOG

OSATS - global rating scale of operative performance

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

	1	2	3	4	5
Respect for Tissue:	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time and Motion:	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency
Knowledge and handling of instrument:	Lack of Knowledge of Instruments		Competent use of instruments but occasionally appeared stiff or awkward		Obvious familiarity with instruments
Flow of operation:	Frequently stopped procedure and seemed unsure of next move		Demonstrated some forward planning with reasonable progression of procedure		Obviously planned course of procedure with effortless flow from one movement to the next
Use of assistants:	Consistently placed assistants poorly or failed to use assistants		Appropriate use of assistants most of the time		Strategically used assistants to the best advantage at all times
Knowledge of specific procedure:	Deficient knowledge. Needed specific instructions at most steps		Knew all important steps of procedure		Demonstrated familiarity with all aspect of operation

Figure 1. OSATS form.

2009] During this implementation study, residents were instructed to register an OSATS assessment of every procedure that they performed as a primary surgeon during their three-month rotation in gynaecological surgery. Procedures during which a resident independently performed some important steps were included as well. After the supervising consultant had filled out the OSATS form, the results were discussed with the resident in order to provide him/her with constructive feedback per domain.

While the assessed trainees were PGY 4 Ob/Gyn residents, the supervisors could be any gynaecologist working as a consultant at the department who was present supervising the surgical procedure. They were instructed how to complete the OSATS form. In essence, the instruction was to mark the number on the Likert-scale corresponding to the resident's performance on each domain, irrespective of the training level.

Individual learning curves

All OSATS were collected, and data were analysed using an SPSS-program for Windows (SPSS version 16.0 SPSS Inc., Chicago, IL). The total score of each OSATS was calculated by adding up the score of the six domains (at minimum 6 and maximally 30 points). An OSATS score of 24 points equals the score in which each domain at average is rated with 4 points (75% of the maximally score that ranges from 1 to 5). This score was chosen as a threshold for good surgical performance, in the absence of benchmark criteria in other studies. Learning curves for each individual resident were drawn by plotting his/her OSATS scores against the total caseload during a clinical rotation, regardless of which procedures were performed. To establish the caseload, all consecutively performed procedures that were assessed with an OSATS were numbered. For each resident, the mean OSATS score during the rotation was calculated, and progression in time was illustrated by mapping a regression line.

Construct validity

No 'gold standard' is available to measure surgical performance. Therefore, the construct validity (i.e. the extent to which a test measures the trait that it purports to measure) should be used to verify the quality of an assessment tool for surgical skills.[Feldman et al., 2004a; Moorthy et al., 2003] In this study, the construct validity of OSATS was established by testing the hypothesis that surgical performance improves as the procedure-specific experience accrues. For that purpose, the average learning curve for the 'average' procedure was mapped by plotting the OSATS score against the procedure-specific caseload. The procedure-specific caseload was also based on the number of assessed procedures.

To test this hypothesis, a linear relation between OSATS score and experience was assumed. The advantage of simplifying the average procedure-specific learning curve to a straight line is that the performance level at the start can be determined, as well as the amount of progression in technical surgical skills, taking individual performance levels and learning potential into account. Therefore, a linear mixed model was fitted as random coefficients model with a random slope and a random intercept per resident. P-values $<.05$ were considered statistically significant, and ninety-five per cent confidence intervals (95% CI) were calculated.

Objectivity of assessment with OSATS

After this implementation study, the opinion of assessed trainees and supervisors was questioned regarding the objectivity of an assessment with an OSATS. They were asked to rate the OSATS on a Likert-scale ranging from 1 "subjective" to 5 "objective". The assessed trainees were residents who were recruited during an education afternoon in the LUMC of which the attendance was obligatory during Ob/Gyn residency training. The supervisors were the same consultants who had participated in the implementation study.

RESULTS

Nine residents attended a three-month clinical rotation in gynaecological surgery, and agreed to participate in the study. Three were male and six were female. Nineteen different types of procedures were assessed with an OSATS, and the total number of procedures was 319. Among these procedures, 39% were abdominal, 31% were laparoscopic, and 20% were procedure with a vaginal approach, and the remaining 10% were hysteroscopies (Table 1). On an individual basis, the median number of procedures assessed was 40 (range 12-60).

Individual learning curves

The nine individual learning curves were drawn by plotting OSATS scores against the total caseload (regardless of which specific procedure had been performed) during the clinical rotation (Figure 2). The regression lines of these curves are displayed too, together with the threshold of 24 (out of 30) OSATS points. Regression analysis revealed that the two residents with the lowest average scores (resident A and B) did not reach the threshold of 24 points within their clinical rotation. Resident C and D reached the threshold while nearing the end of their rotation. Only resident H and I achieved relatively high scores at the start of the three-month period and continued to show improvement.

Average procedure-specific learning curve

Additionally, the average OSATS scores were plotted against the experience, i.e. the procedure-specific caseload, for the first ten procedures (Figure 3). The resulting average learning curve within procedure passed the threshold of 24 points at a caseload of five procedures. Additionally, a plateau in performance was reached after a caseload of eight procedures. To establish the construct validity of OSATS it was tested whether the OSATS score increased significantly with an increasing caseload using a linear mixed model. The slope of the general learning curve was 1.10 OSATS points per assessed procedure ($p<.01$, 95% CI: 0.44 – 1.77). In other words, the average performance based on total OSATS score improved by 1.10 points for every consecutively performed procedure.

An OSATS score of 24 was set as the performance standard. The dotted line is based on linear mixed model analysis.

Objectivity of the assessment

The supervisors were 21 gynaecologists, all working as consultants at the Department of Gynaecology at the LUMC. The median OSATS score given to residents by each supervisor

Table 1.

Procedures	Number assessed with an OSATS
Laparoscopic procedures	98
Diagnostic laparoscopy or sterilization	23
(Bilateral) Salpingo-oophorectomy	41
Cystectomy	17
Ectopic Pregnancy (tobotomy or tubectomy)	4
Total laparoscopic hysterectomy	13
Hysteroscopic procedures	31
Diagnostic hysteroscopy	12
Therapeutic hysteroscopy	19
Abdominal procedures	125
Abdominal hysterectomy (with (B)SO)	42
Resection myoma, endometrioma or adnexectomy	6
Caesarean section	64
(Interval) debulking	7
Sacrocolpopexy	5
Procedure with a vaginal approach	65
Vaginal hysterectomy	43
Anterior and/or posterior colporrhaphy	6
(Partial) vulvectomy	2
Operation of cervix (cerclage or conization)	8
Anal sphincter repair	4
Laser treatment vulva	2
Labioplasty	1
Total	319

ranged from 18 to 30, and the number of assessed procedures ranged from 1 to 114. Moreover, some gynaecologists assessed only one specific procedure (e.g., a caesarean section), while others assessed the entire surgical spectrum.

All 24 residents who were present at the obligatory education afternoon answered the question about the OSATS. One person was excluded from analysis due to inexperience with this assessment form because of being just at the start of residency training. Residents rated the OSATS with a median score of 2 (range 1-4 on a 5-point Likert scale with 1:subjective to 5:objective). The median score of the supervisors was 3 (range 1-4).

DISCUSSION

Intraoperative OSATS can be used to assess resident's surgical training over time. By plotting the OSATS score against experience it can be determined whether, and how much, progression is

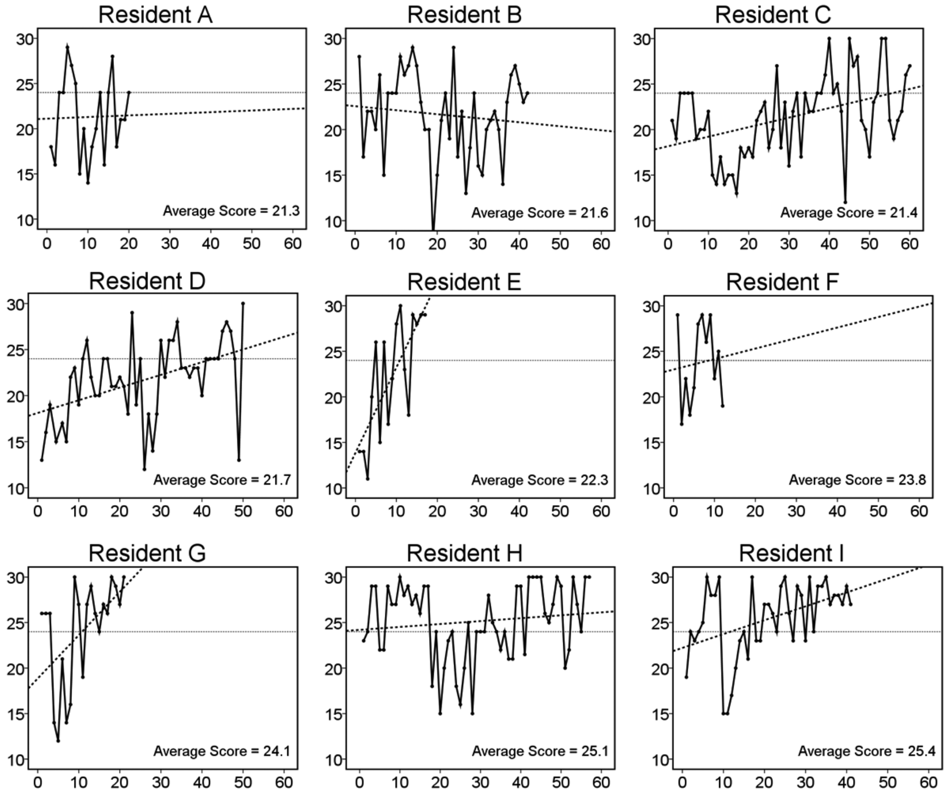


Figure 2. Individual learning curves (regardless the type of procedure performed). x = total caseload expressed in number of assessed procedure (regardless of the type of procedure performed), y = performance expressed in total OSATS score, dotted line = individual regression line.

present. The use of an objective assessment tool is a new way to establish learning curves. Prior parameters are the operation time, the complication rate and the conversion rate in case of laparoscopic procedures.[Altgassen et al., 2004; Kolkman et al., 2007a] However, operation time and complication rate have shown to be crude and indirect as these largely depend on the difficulty of the individual surgical case (e.g. the co-morbidity of a patient), and the supervising surgeon.[Moorthy et al., 2003] The intraoperative use of OSATS may overcome these disadvantages.

Two out of nine residents did not progress beyond the benchmark level of 24/30 OSATS points within the three-month clinical rotation. This failure is likely to be a sign of stagnation of their learning process, and can only partially be explained by the coincidence that they encountered more complex procedures later in their rotation. Additionally, only two residents showed good performance during the entire clinical rotation, taking the average OSATS scores and the progression into account. This small proportion illustrates the concern whether current residency programs with work hour restrictions sufficiently fulfil the need to master surgical proficiency.

Secondly, the construct validity of the OSATS for assessment purposes was revealed by confirming the hypothesis (i.e. the construct) to be true that a resident's OSATS' score

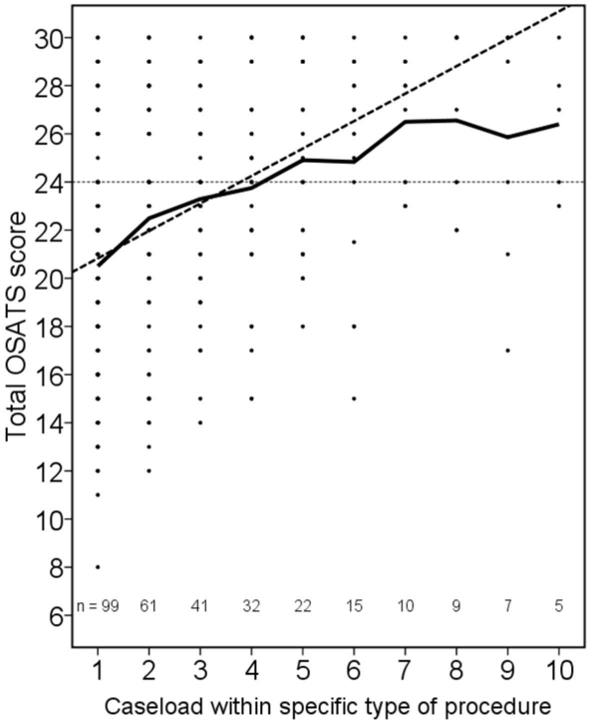


Figure 3. Average Objective Structured Assessment of Technical Skills (OSATS) scores plotted against procedure-specific learning curve for the first 10 procedures.

improves as procedure-specific experience accrues. This is not the conventional way to prove the construct validity. However, it is a more subtle approach than the often used method to confirm the ability of an assessment tool to discriminate between two groups of hugely varying level of experience. That was done by Aggarwal et al. who revealed that experienced surgeons have higher OSATS scores than novice surgeons for one standardized procedure, the laparoscopic cholecystectomy.[Aggarwal et al., 2008] The straight line model we used as an argument for the construct validity has two limitations. Surgical performance cannot infinitely improve (the maximum OSATS score is 30 points), and secondly, the learning curve for surgical skills consists of an initial steep phase, then changes slowly until the curve becomes more flat.[Dagash et al., 2003] However, the advantage of simplifying resident's learning curve to a straight line, and additional analysis with linear mixed model, is that progression in surgical skills can be quantified taking the individual level of performance and learning potential into account. From this data, it was found that a resident's performance improves with 1.14 OSATS points at average every time the same procedure is performed (and assessed). Of course, we may not simply generalize this conclusion, because this increase is the average of 19 very different surgical procedures.

The aforementioned formation of a plateau in the real situation is observed in the average procedure-specific learning curve. This plateau is achieved after a caseload of eight (of the same) procedures. This is in accordance with results of a questionnaire held among residents is which they judged a number of ten of the same procedures necessary to be a safe and confident surgeon.[Rattner et al., 2001] Again, this value of this generalization is limited because of the heterogeneous range of assessed procedures.

This study was conducted under regular clinical conditions. Therefore, even the same procedures widely varied with respect to difficulty and complication risk. Also, variation shall have been present in the extent to which consultants allowed residents to independently perform a surgical procedure. Furthermore, the assessment rate might not be 100 per cent. The resulting selection bias may be in favour of the best performed procedures. However, not all procedures need to be assessed to gain insight in the progression of an individual resident. More importantly, the intended objectivity of assessment with an OSATS seems to be disappointing, taking the finding that none of the residents, nor any staff member, valued the OSATS to be objective into account. Additionally, the number of assessed procedures and the OSATS-score varied enormously among the consultants. This variation occurred despite the uniform instruction that all supervisors had received. An attempt to achieve more uniformity might be realized by organizing additional training for the supervisors in the registration of an OSATS. However, in our opinion, the effect of such training is limited. No information can be added to the original instruction to mark the number on the rating scale corresponding to the resident's performance on each domain, irrespective of the training level. Moreover, an assessment based on the opinion of an individual will never be free from subjectivity. A study in which residents all perform at least ten of the same procedures consecutively would have allowed firmer conclusions about the learning for curve of that specific procedure. However, insight in daily practice is obtained by analysing the heterogeneous data of our study, and illustrates the study's relevance.

In conclusion, assessment with OSATS during residency has many advantages. OSATS-based learning curves have the potential to select residents in need of more guidance during their learning process. Consequently, cues are provided to tailor surgical skills training to individual needs. An OSATS does not need to concern the entire procedure; (small) steps of the procedure can be evaluated as well. Additionally, it provides a framework of structured instantaneous feedback on surgical skills in general (total OSATS score). Theoretically, the specific domains of technical skills (e.g. respect for tissue, knowledge and handling of instruments) also provide cues for identifying individual needs. However, the information that the domain-specific scores add is limited as revealed from the small variety of score within one OSATS. Ideally, the structural feedback on surgical performance using assessment with OSATS will enhance the efficiency of the spare learning moments in the OR. From that point of view, we consider the general global rating scale of OSATS to be suitable for large scale implementation in the OR.

However, the inherent subjectivity of an assessment using an opinion based tool needs to be taken into account. Regarding the results of the questionnaire and the enormous variation in supervisor's scores, an OSATS unfortunately is not as objective as it intended to be. This is an important limitation of the OSATS that, to our knowledge, has not been highlighted in other publications about this assessment tool. Furthermore, there are other ways to evaluate a subject's surgical skills. Therefore, caution needs to be exercised in using OSATS for certification and qualification purposes, or in advising an individual resident to choose for a non-surgical specialization if the OSATS-based performance continues to be disappointing. Though, presently, it seems to be the best tool available.