Chapter 8

Objective versus subjective assessment of laparoscopic skill:
OSATS vs TrEndo

Pieter J. van Empel, M.D., Lennart B. van Rijssen, Joris P. Commandeur, M.D., Mathilde G.E. Verdam, MSc., Judith A. Huirne, M.D., Ph.D., Fedde Scheele, M.D., Ph.D., H. Jaap Bonjer, M.D., Ph.D., W. Jeroen Meijerink, M.D. Ph.D.

Minimally Invasive Surgery 2013, Accepted for publication
Abstract

Background

The equality of subjective- and objective-assessment methods in laparoscopic surgery are unknown. The aim of this study was to compare a subjective assessment method to an objective assessment method to evaluate laparoscopic skill.

Methods

A prospective observational cohort study was conducted. 72 residents completed a basic laparoscopic suturing task on a box trainer at two consecutive assessment points. Laparoscopic skill was rated subjectively using the Objective Structured Assessment of Technical Skills (OSATS) list, and objectively using the TrEndo augmented-reality simulator.

Results

TrEndo scores between the two assessment points correlated. OSATS scores did not correlate between the two assessment points. There was a correlation between TrEndo and OSATS scores at the first assessment point, but not at the second assessment point. Overall, OSATS scores correlated with TrEndo scores. There was a greater spread within OSATS scores compared to TrEndo scores.

Conclusion

OSATS scores correlated with TrEndo scores. The TrEndo may be more responsive at rating an individual's laparoscopic skill, as demonstrated by a smaller overall
spread in TrEndo scores. The additional value of objective assessment methods over conventional assessment methods as provided by laparoscopic simulators should be investigated.
Introduction

Simulation based training in laparoscopic or minimally invasive surgery (MIS) is gaining recognition. To enhance patient safety the initial learning curve in MIS may be moved outside the operating room (OR). Several validation studies demonstrated simulator based training to improve laparoscopic psychomotor skills and OR performance\textsuperscript{1-3}. Simulation based training is also cost-effective and easily implementable in a surgical training curriculum\textsuperscript{4,5}.

Three classes of MIS simulators are currently available: computerized virtual reality trainers (VR), traditional box trainers (BT) and a combination of the former two; augmented reality trainers (AR)\textsuperscript{6,7}. All simulators aim for a maximal realistic setting. However, laparoscopic basic skills training remains unrealistic as VR and AR simulators are computer based. Furthermore, VR and AR simulators do not provide tactile (haptic) feedback. Box trainers do preserve a realistic setting including haptic feedback and augmented depth perception\textsuperscript{8}, but would lack objective assessment methods and a low fidelity to actual laparoscopic procedures\textsuperscript{9}.

Training in MIS should include objective assessment. Various MIS assessment methods have been developed and typically include checklists for various task-specific components based on a supervisor's evaluation, however such checklists are indicative mostly of procedural performance rather than a measure of technical ability\textsuperscript{10-12}. A widely validated MIS assessment method is the ‘Objective Structured Assessment of Technical Skill’ (OSATS) list\textsuperscript{13,14}. The OSATS is a procedure-specific checklist on a global rating scale. The OSATS is currently regarded as the gold standard for assessment of MIS skill. An OSATS score is rapid and functional in both
real-time performance in the OR and in a simulated training environment\textsuperscript{15}. However, as performance is evaluated by a supervisor, the OSATS are subjective and subject to inter-observer bias\textsuperscript{14}. Furthermore, only one study established a score cut-off point on the OSATS to classify a trainee as being proficient in autonomously performing a laparoscopic procedure\textsuperscript{16;17}.

Objective classification of MIS skill is often based on parameters recorded by MIS simulators such as the amount of time taken to complete a task, knot tension, or motion analysis parameters (MAPS) such as instrument path length. Various validation studies have demonstrated objective classification of MIS skill using laparoscopic simulators although there is little evidence that these systems can assess a residents' individual laparoscopic performance correctly\textsuperscript{8}.

One such simulator, The TrEndo (Delft University of Technology, Delft, the Netherlands), was previously validated by Chmarra et al. and by our research group\textsuperscript{18;19}. The TrEndo records various MAPS as explained below. The TrEndo is a relatively new augmented reality simulator situated on a laparoscopic box trainer. The TrEndo incorporates the advantages of a computer-based simulator while preserving haptic feedback and a 'realistic' environment as provided by the box trainer.

With the introduction and rapidly increasing number of objective assessment tools in laparoscopic surgery, the equality of subjective assessment methods to objective assessment methods to rate laparoscopic skill are relatively unknown. The aim of this study was to compare a subjective assessment tool (the OSATS) to an objective assessment tool (the TrEndo) in the evaluation of laparoscopic skill.
Materials & Methods

A prospective observational cohort study was conducted in The Netherlands and Belgium between February 1st and November 31st 2010.

Participants

Residents in urology, gynecology and surgery attending a laparoscopic suturing course organized by the VU University Medical Center at hospitals in The Netherlands and Belgium voluntarily enrolled in this study. Course participation was allowed when residents had completed basic open- and laparoscopic- training programs. The laparoscopic suturing course comprises two training days, in-between which a six week autonomous training period is located.

Task

Participants completed a standardised laparoscopic square knot in a laparoscopic box trainer (as described below) on an artificial skin patch using a 15 cm single 3-0 silk suture. A 5-minute time limit was set. The task was assessed at the start of the first training day (pre-assessment), at the end of the first training day (post-assessment), and after the 6-week autonomous training period (follow-up assessment) using both the OSATS and the TrEndo assessment methods simultaneously (also described below). The objective of the first (pre-) assessment point was solely to gain acquaintance to equipment- and assessment- methods used. Equipment and instruments used were kept identical at all times.
OSATS

The OSATS list contains 9 task-specific components, evaluated by a supervisor (in our case a senior surgeon). Regarding a laparoscopic square knot, task-specific components include various technical components, instrument configuration, tissue fixation, needle positioning and flow of movements. Each component is given a score between 1 and 5, to generate an overall score between 5 and 45.

Laparoscopic box trainer & TrEndo

Laparoscopic box trainers (Camtronics Nederland B.V., Son, The Netherlands) simulate an abdominal cavity using an aluminium frame and allow regular insertion of traditional trocarts with conventional laparoscopic instruments (B. Braun Medical B.V., Oss, Holland) and a camera connected to a video monitor on which the simulated environment is viewed.

The TrEndo is constructed as a trocart on a laparoscopic training box through which laparoscopic instruments may be regularly inserted allowing free instrument motions (figure 1). Five motion analysis parameters (MAPs) are recorded individually for the right and left hand including path length ($mm$, length of the curve as described by tip of the instrument), insertion distance ($mm$, total distance traveled by the instrument along its axis), angular area ($deg^2$, relationship of the distance between the outermost positions of the instrument), volume ($mm^3$, thee-dimensional space used), and time taken to complete the task ($sec^{18}$).
Comparison of OSATS (subjective) versus TrEndo (objective)

In our previous work, we demonstrated construct validity of the TrEndo. The results (based on a logistic regression formula) were used to generate a TrEndo score between 0 and 10 in this study, with a higher score indicating superior performance. OSATS scores were divided by 4.5 to also yield a score between 0 and 10.

We only compared evaluations of laparoscopic skill between assessment points 2 (post) and 3 (follow-up) as participants were often unfamiliar with used materials, devices and assessment protocols, and as some residents were unable to perform a laparoscopic suture at the start of the suturing course. Evaluation of the first assessment point would therefore be meaningless. It was not the intention of this study to investigate a learning curve.

Statistical Analysis
Statistical analysis was performed using SPSS 17.0 for Windows (SPSS Inc., Chicago, Illinois, USA). A nominal significance level of 0.05 was used. All tests were performed two sided. Values for continuous variables are given as mean (SD). We calculated mean TrEndo and OSATS scores and compared changes in TrEndo and OSATS scores over the two assessment points. We also analyzed correlations between the difference in TrEndo and the difference in OSATS scores between assessment point two and three. Possible differences between specialisms were analyzed. We used an ICC instead of Pearson correlation, to account for inter-individual variance. Mean (+-SD) OSATS and TrEndo scores were calculated. The assumption of normality was checked by inspecting skewness and kurtosis values and histograms and q-q plots of all variables. OSATS scores appear to follow a normal distribution, but TrEndo scores are somewhat negatively skewed. To account for this deviation of normality non-parametric alternatives when used when available.

To determine if a change in TrEndo or OSATS score was significant, wilcoxon signed rank test was used. A repeated measures ANOVA was used to compare changes in TrEndo and OSATS scores over the two assessment points. Spearman correlation was used to analyze correlations between TrEndo and OSATS scores for both assessment points, and the change in TrEndo and the change in OSATS scores between assessment point two and three. Furthermore, an overall intra-class correlation (ICC) was calculated, in addition to ICCs for both TrEndo and OSATS separately.

Results
In total, 72 participants voluntarily participated in this study. Out of 72, 66 (91.7%) participants completed both assessment points on the TrEndo and 71 (98.6%) participants completed both OSATS assessments. Six residents (8.3%) were unable to attend all assessment points due to logistic reasons, however average scores were similar compared to residents attending each assessment point. Most residents were active in general surgery (n=47; 65.3%), gynecology (n=15, 20.8%) and urology (n=10, 13.8%). All residents had completed two years of residency. At post-assessment, the mean OSATS score was 7.4 (1.3 SD) and the mean TrEndo score was 6.5 (3.9 SD). At follow-up the mean OSATS score was 7.5 (1.3 SD) and the mean TrEndo score was 7.6 (3.4 SD). OSATS scores did not increase; difference was 0.1 (p=0.365), while TrEndo score increased by 1.1 (p=0.006, figure 2).

In total, an increase in OSATS score correlated with an increase in TrEndo score (r=.342, p=0.005, n=66). However, the increase in TrEndo scores between assessment points was significantly larger than the increase in OSATS scores (sphericity assumed p = 0.046, n=66).

<table>
<thead>
<tr>
<th></th>
<th>OSATS follow-up</th>
<th>TrEndo post</th>
<th>TrEndo follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSATS post</td>
<td>.05</td>
<td>.42</td>
<td>.13</td>
</tr>
<tr>
<td>OSATS follow-up</td>
<td></td>
<td>.09</td>
<td>.08</td>
</tr>
<tr>
<td>TrEndo post</td>
<td></td>
<td></td>
<td>.49</td>
</tr>
</tbody>
</table>

Table 1: Spearman correlations between OSATS and TrEndo scores for post- and follow-up assessment.
For all four variables ICC was .52 (95% CI: .30 ; .68), while ICC for TrEndo scores across the assessment points was .60 (95% CI: .35 ; .76) and ICC for OSATS scores across the assessment points was .17 (95% CI: -.32 ; .48).

Using ICC analysis, OSATS scores between post and follow-up assessment did not correlate, whereas TrEndo scores between post and follow-up assessment did correlate. There was a correlation between the two assessment methods (TrEndo-OSATS) at the post assessment. However, at the follow-up assessment TrEndo and OSATS no longer correlated. Overall, there was no significant difference in OSATS or TrEndo scores between residents of different specializations (p= 0,414).

Figure 2: Mean scores and standard errors at two assessment points. OSATS: Objective Assessment of Surgical Techniques.

**Discussion**
Objective assessment methods are gaining in popularity compared to subjective assessment methods in laparoscopic surgery. Yet, no best evidence based assessment method is currently available. In this study we aimed to evaluate equality between a subjective- and objective-assessment of a basic laparoscopic suturing task.

In a large and heterogeneous study group of 72 residents who were followed over 6 weeks and assessed at the start and end of this period, we found a significant increase in TrEndo scores, but a non-significant increase in OSATS scores. An increase in OSATS scores correlated with an increase in TrEndo scores. The average increase in TrEndo score was greater than the increase in OSATS score. We feel this is not due to an inappropriate interpretation of laparoscopic skill by either assessment method, as overall TrEndo outcomes globally correlated with OSATS scores. The discrepancy may be explained by the difference in outcome parameters used, as the OSATS rating scale is based on task performance, whereas the TrEndo is indicative of task efficiency (and consequently may also be indicative of laparoscopic performance, as discussed in our previous manuscript). It is the author’s belief that improvement of task efficiency precedes improved performance during training, as could be concluded from our study results.

At post assessment OSATS and TrEndo scores correlated, but did not correlate at follow up assessment. We assume this is due to the subjectivity of the OSATS, as TrEndo scores remain correlated with themself.

The spread in OSATS scores was smaller than the spread in TrEndo scores. As the OSATS are scaled from 0 to 5, examiners tend to reward an average score, e.g. 3 or
4. A score of 0 or 5 is rarely given. Therefore the TrEndo might be more distinctive than the OSATS in scoring laparoscopic skill.

Laparoscopic suturing is a unique simulator-based task as it incorporates all basic laparoscopic skills including ambidexterity, depth perception, material and instrument handling and instrument manipulation. Therefore proficiency in a laparoscopic suturing task is a valuable indication of laparoscopic skill and a useful benchmark in the assessment of laparoscopic skill\(^20\).

To our knowledge, only two previous studies compared the OSATS to a validated ‘objective’ assessment method of MIS skill. Both of these studies demonstrated adequate correlations between the 2 models of assessment. The Imperial College Surgical Assessment Device (ICSAD) – which also implements various task efficiency parameters – was compared to the OSATS to establish construct and concurrent validity for a MIS simulator\(^21\), and another study compared the ICSAD and a video assessment to the OSATS\(^22\).

Our study contained several limitations. First, we did not perform a power calculation. However, compared to other comparable studies our study contained a substantial study group. Second, our study group consisted of voluntary participants attending the laparoscopic suturing course with consequent possible selection bias and for that reason it was not possible to generalize our results. Third, the OSATS were rated by various faculty members who were not blinded, therefore inter-observer bias is not excluded. Fourth, we did not divide residents by specialism. Residents from various specialties may not have equal laparoscopic experience. However, all residents had completed a basic laparoscopic surgical skills program during the first year of residency. Fifth, we did not weigh individual components within the OSATS or
TrEndo scores. The weight of various task-specific components is however not known and subject to evaluation\(^\text{19}\).

Adequate training in laparoscopic surgery should include an objective assessment means. Superiority of ‘objective’ assessment methods over gold standard subjective evaluation such as by the OSATS remains unclear. It is the author’s expectation that the TrEndo will be more responsive than the OSATS to rate laparoscopic skill, as a greater variation is possible within recorded MAPs compared to the OSATS. A greater possible variation within the rating scale and the absence of inter-observer bias renders the TrEndo more effective at recording individual progress. Task efficiency has already been shown to correlate to OR-performance\(^\text{22}\). Future studies should investigate the added value of objective assessment parameters as provided by MIS simulators to standard subjective evaluation such as by the OSATS. An objective assessment of laparoscopic skill should be integrated into an extensive validated training program in MIS.

\textbf{CONCLUSION}

Most laparoscopic training programs are not evidence-based and there is no global consent on an objective assessment of MIS skill\(^\text{23}\). Current assessment relies on subjective evaluation based on a supervisor’s recollection\(^\text{24,25}\). In this study we demonstrated that an objective assessment method of a basic laparoscopic task by
means of the TrEndo laparoscopic simulator globally correlated with the subjective gold standard by means of the OSATS. The TrEndo might be more responsive than the OSATS in scoring laparoscopic skill, and might be more effective at recording individual progress. Future studies should investigate the added value of objective assessment methods over subjective assessment methods in laparoscopic training programs.
Reference List


