

## Research Article

# Objective versus Subjective Assessment of Laparoscopic Skill

Pieter J. van Empel,<sup>1</sup> Lennart B. van Rijssen,<sup>1</sup>  
Joris P. Commandeur,<sup>1</sup> Mathilde G. E. Verdam,<sup>2</sup> Judith A. Huirne,<sup>3</sup>  
Fedde Scheele,<sup>4</sup> H. Jaap Bonjer,<sup>1</sup> and W. Jeroen Meijerink<sup>1</sup>

<sup>1</sup> Department of Surgery, VU University Medical Centre, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

<sup>2</sup> Department of Medical Psychology, University of Amsterdam, P.O. Box 19268, 1000 GG Amsterdam, The Netherlands

<sup>3</sup> Department of Obstetrics and Gynaecology, VU University Medical Centre, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

<sup>4</sup> Department of Obstetrics and Gynaecology, Sint Lucas Andreas Hospital, P.O. Box 9243, 1006 AE Amsterdam, The Netherlands

Correspondence should be addressed to Pieter J. van Empel; [p.vanempel@vumc.nl](mailto:p.vanempel@vumc.nl)

Received 25 May 2013; Accepted 13 June 2013

Academic Editors: F. Agresta, M. Barczyński, and Y. Izci

Copyright © 2013 Pieter J. van Empel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**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. Seventy-two 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, an 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 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.

## 1. 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 [1–3]. Simulation-based training is also cost-effective and easily implementable in a surgical training curriculum [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) [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 [8] but would lack objective assessment methods and a low fidelity to actual laparoscopic procedures [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 supervisor's evaluation; however, such checklists are indicative mostly of procedural performance rather than a measure of technical ability [10–12]. A widely validated MIS assessment method is the "Objective Structured Assessment of Technical Skill" (OSATS) list [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 [15]. However, as performance is evaluated by a supervisor, the OSATS are subjective and subject to interobserver bias [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 [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 residents' individual laparoscopic performance correctly [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 [18, 19]. The TrEndo records various MAPS as explained in the following. 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 is 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.

## 2. Materials and Methods

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

**2.1. 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 [7]. The laparoscopic suturing course comprises two training days, in-between which a six-week autonomous training period is located.

**2.2. Task.** Participants completed a standardised laparoscopic square knot in a laparoscopic box trainer (as described in the following) 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 (preassessment), at the end of the first training day (postassessment), and after the 6-week autonomous training period (follow-up assessment) using both the OSATS and the TrEndo assessment methods simultaneously (also described in the following).



FIGURE 1: The TrEndo tracking device.

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.

**2.3. 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.

**2.4. Laparoscopic Box Trainer and 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 trocars 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 trocar 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 hands 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 ( $\text{deg}^2$ , relationship of the distance between the outermost positions of the instrument), volume ( $\text{mm}^3$ , three-dimensional space used), and time taken to complete the task (sec) [18].

**2.5. Comparison of OSATS (Subjective) versus TrEndo (Objective).** In our previous study 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.

**2.6. Statistical Analysis.** Statistical analysis was performed using SPSS 17.0 for Windows (SPSS Inc., Chicago, IL, 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 points two and three. Possible differences between specialisms were analyzed. We used an ICC instead of Pearson correlation, to account for interindividual variance. Mean ( $\pm$ 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 nonparametric alternatives were 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 points two and three. Furthermore, an overall intraclass correlation (ICC) was calculated, in addition to ICCs for both TrEndo and OSATS separately.

### 3. 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

TABLE 1: Spearman correlations between OSATS and TrEndo scores for post- and follow-up assessments.

	OSATS follow-up	TrEndo post-assessment	TrEndo follow-up assessment
OSATS post-assesment	0.05	0.42	0.13
OSATS follow-up assessment		0.09	0.08
TrEndo post-assesment			0.49

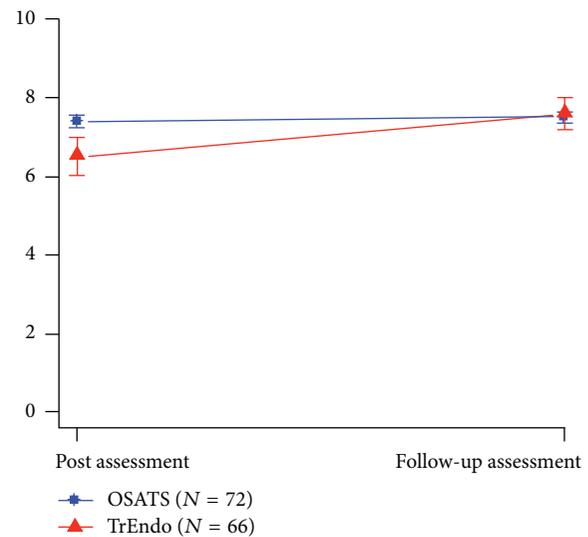


FIGURE 2: Mean scores and standard errors at two assessment points. OSATS: Objective Assessment of Surgical Skills.

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 ( $r = 0.342$ ,  $P = 0.005$ ,  $n = 66$ ). However, the increase in TrEndo scores between assessment points was significantly larger than the increase in OSATS scores (sphericity assumed  $P = 0.046$ ,  $n = 66$ ).

For all four variables ICC was 0.52 (95% CI: 0.30; 0.68), while ICC for TrEndo scores across the assessment points was 0.60 (95% CI: 0.35; 0.76), and ICC for OSATS scores across the assessment points was 0.17 (95% CI: -0.32; 0.48).

Using ICC analysis, OSATS scores between post- and follow-up assessments did not correlate, whereas TrEndo scores between post- and follow-up assessments did correlate. There was a correlation between the two assessment methods (TrEndo-OSATS) at the postassessment. However, at the follow-up assessment TrEndo and OSATS no longer correlated (Table 1). Overall, there was no significant difference in OSATS or TrEndo scores between residents of different specializations ( $P = 0.414$ ).

## 4. 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 nonsignificant 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 that 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 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 that this is due to the subjectivity of the OSATS, as TrEndo scores remain correlated with themselves.

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, for example, 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 interobserver 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 [19].

Adequate training in laparoscopic surgery should include an objective assessment means. Superiority of "objective" assessment methods over gold standard subjective evaluation such as the OSATS remains unclear. It is 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 interobserver bias renders the TrEndo more effective at recording individual progress. Task efficiency has already been shown to correlate to OR-performance [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.

## 5. Conclusion

Most laparoscopic training programs are not evidence-based, and there is no global consent on an objective assessment of MIS skill [23]. Current assessment relies on subjective evaluation based on supervisor's recollection [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.

## Conflict of Interests

Pieter J. van Empel, Lennart B. van Rijssen, M.D., Joris P. Commandeur, M.D., Mathilde G. E. Verdam, M.S., Judith Huirne, M.D., Ph.D., Fedde Scheele, M.D., Ph.D., H. Jaap Bonjer, M.D., Ph.D., and Wilhelmus J. Meijerink, M.D., Ph.D., have no conflict of interests or financial ties to disclose related to this paper.

## Acknowledgments

The authors thank Dr. J. J. van den Dobbelsteen, Department of Biomechanical Engineering, Delft University of Technology, for his contribution and technical support regarding the TrEndo tracking device. The authors want to thank

Mr. R. P. M. de Hoon, Department of Surgery, VU University Medical Center, for his important contributions to the organization of the Advanced Suturing Course.

## References

- [1] G. M. Fried, L. S. Feldman, M. C. Vassiliou et al., "Proving the value of simulation in laparoscopic surgery," *Annals of Surgery*, vol. 240, no. 3, pp. 518–528, 2004.
- [2] D. J. Scott, P. C. Bergen, R. V. Rege et al., "Laparoscopic training on bench models: better and more cost effective than operating room experience?" *Journal of the American College of Surgeons*, vol. 191, no. 3, pp. 272–283, 2000.
- [3] N. E. Seymour, A. G. Gallagher, S. A. Roman et al., "Virtual reality training improves operating room performance results of a randomized, double-blinded study," *Annals of Surgery*, vol. 236, no. 4, pp. 458–464, 2002.
- [4] M. T. Goova, L. A. Hollett, S. T. Tesfay et al., "Implementation, construct validity, and benefit of a proficiency-based knot-tying and suturing curriculum," *Journal of Surgical Education*, vol. 65, no. 4, pp. 309–315, 2008.
- [5] D. J. Scott, E. M. Ritter, S. T. Tesfay, E. A. Pimentel, A. Nagji, and G. M. Fried, "Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training," *Surgical Endoscopy and Other Interventional Techniques*, vol. 22, no. 8, pp. 1887–1893, 2008.
- [6] S. M. B. I. Botden and J. J. Jakimowicz, "What is going on in augmented reality simulation in laparoscopic surgery?" *Surgical Endoscopy and Other Interventional Techniques*, vol. 23, no. 8, pp. 1693–1700, 2009.
- [7] E. J. B. Furnée, P. J. Van Empel, B. M. Delavary, D. L. Van Der Peet, M. A. Cuesta, and W. J. H. J. Meijerink, "Evaluation of a technical skills training program in surgical residents," *Journal of Laparoendoscopic and Advanced Surgical Techniques*, vol. 19, no. 5, pp. 615–621, 2009.
- [8] P. J. Van Empel, W. M. Van Der Veer, L. B. Van Rijssen et al., "Mapping the maze of minimally invasive surgery simulators," *Journal of Laparoendoscopic and Advanced Surgical Techniques*, vol. 22, no. 1, pp. 51–60, 2012.
- [9] R. Aggarwal, K. Moorthy, and A. Darzi, "Laparoscopic skills training and assessment," *British Journal of Surgery*, vol. 91, no. 12, pp. 1549–1558, 2004.
- [10] S. Cotin, N. Stylopoulos, M. Ottensmeyer, P. Neumann, D. Rattner, and S. Dawson, "Metrics for laparoscopic skills trainers: the weakest link!" in *Medical Image Computing and Computer-Assisted Intervention*, vol. 2488 of *Lecture Notes in Computer Science*, pp. 35–43, 2002.
- [11] K. Moorthy, Y. Munz, A. Dosis, F. Bello, and A. Darzi, "Motion analysis in the training and assessment of minimally invasive surgery," *Minimally Invasive Therapy and Allied Technologies*, vol. 12, no. 3–4, pp. 137–142, 2003.
- [12] R. K. Reznick, "Teaching and testing technical skills," *The American Journal of Surgery*, vol. 165, no. 3, pp. 358–361, 1993.
- [13] D. Dath, G. Regehr, D. Birch et al., "Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills," *Surgical Endoscopy and Other Interventional Techniques*, vol. 18, no. 12, pp. 1800–1804, 2004.
- [14] J. A. Martin, G. Regehr, R. Reznick et al., "Objective structured assessment of technical skill (OSATS) for surgical residents," *British Journal of Surgery*, vol. 84, no. 2, pp. 273–278, 1997.
- [15] B. A. Goff, G. M. Lentz, D. Lee, D. Fenner, J. Morris, and L. S. Mandel, "Development of a bench station objective structured assessment of technical skills," *Obstetrics & Gynecology*, vol. 98, no. 3, pp. 412–416, 2001.
- [16] P. D. Van Hove, G. J. M. Tuijthof, E. G. G. Verdaasdonk, L. P. S. Stassen, and J. Dankelman, "Objective assessment of technical surgical skills," *British Journal of Surgery*, vol. 97, no. 7, pp. 972–987, 2010.
- [17] C. B. M. Bijen, J. M. Briët, G. H. de Bock, H. J. G. Arts, J. A. Bergsma-Kadijk, and M. J. E. Mourits, "Total laparoscopic hysterectomy versus abdominal hysterectomy in the treatment of patients with early stage endometrial cancer: a randomized multi center study," *BMC Cancer*, vol. 9, article 23, 2009.
- [18] M. K. Chmarra, S. Klein, J. C. F. De Winter, F. W. Jansen, and J. Dankelman, "Objective classification of residents based on their psychomotor laparoscopic skills," *Surgical Endoscopy and Other Interventional Techniques*, vol. 24, no. 5, pp. 1031–1039, 2010.
- [19] P. J. van Empel, L. B. van Rijssen, J. P. Commandeur et al., "Validation of a new box trainer-related tracking device: the TrEndo," *Surgical Endoscopy*, vol. 26, no. 8, pp. 2346–2352, 2012.
- [20] R. Aggarwal, J. Hance, S. Undre et al., "Training junior operative residents in laparoscopic suturing skills is feasible and efficacious," *Surgery*, vol. 139, no. 6, pp. 729–734, 2006.
- [21] V. Datta, A. Chang, S. Mackay, and A. Darzi, "The relationship between motion analysis and surgical technical assessments," *American Journal of Surgery*, vol. 184, no. 1, pp. 70–73, 2002.
- [22] R. Aggarwal, T. Grantcharov, K. Moorthy et al., "An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room," *Annals of Surgery*, vol. 245, no. 6, pp. 992–999, 2007.
- [23] S. G. C. Kroeze, E. K. Mayer, S. Chopra, R. Aggarwal, A. Darzi, and A. Patel, "Assessment of laparoscopic suturing skills of urology residents: a pan-European study," *European Urology*, vol. 56, no. 5, pp. 865–873, 2009.
- [24] W. Kolkman, R. Wolterbeek, and F. W. Jansen, "Gynecological laparoscopy in residency training program: Dutch perspectives," *Surgical Endoscopy and Other Interventional Techniques*, vol. 19, no. 11, pp. 1498–1502, 2005.
- [25] L. P. Mandel, G. M. Lentz, and B. A. Goff, "Teaching and evaluating surgical skills," *Obstetrics and Gynecology*, vol. 95, no. 5, pp. 783–785, 2000.



**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

