



**HAL**  
open science

## Performance Assessment for Total Laparoscopic Hysterectomy in the Operating Room: Validity Evidence of a Procedure-specific Rating Scale

Patrice Crochet, Antoine Netter, Andy Schmitt, Anna Garofalo, Anderson Loundou, Sophie Knight, Benoit Rabischong, Aubert Agostini

### ► To cite this version:

Patrice Crochet, Antoine Netter, Andy Schmitt, Anna Garofalo, Anderson Loundou, et al.. Performance Assessment for Total Laparoscopic Hysterectomy in the Operating Room: Validity Evidence of a Procedure-specific Rating Scale. *Journal of Minimally Invasive Gynecology*, 2021, 28 (10), pp.1743-1750.e3. 10.1016/j.jmig.2021.02.013 . hal-03525451

**HAL Id: hal-03525451**

**<https://hal.science/hal-03525451v1>**

Submitted on 13 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Performance Assessment for Total Laparoscopic Hysterectomy in the Operating Room: Validity Evidence of a Procedure-specific Rating Scale

Patrice Crochet, PhD, MD, Antoine Netter, MD, Andy Schmitt, MD, Anna Garofalo, MD, Anderson Loundou, PhD, Sophie Knight, MD, Benoit Rabischong, PhD, MD, and Aubert Agostini, MD

*From the Department of Obstetrics and Gynecology, La Conception Hospital, Aix Marseille University (Drs. Crochet, Netter, Schmitt, Garofalo, Knight, and Agostini), Institut Méditerranéen de Biodiversité et d'Écologie marine et continentale, Aix Marseille University, CNRS, IRD, Avignon University (Dr. Netter), Support Unit for Clinical Research and Economic Evaluation, Assistance Publique Hôpitaux de Marseille, Aix Marseille University (Dr. Loundou), Marseille, Department of Obstetrics and Gynecology, Arnaud de Villeneuve Hospital, University of Montpellier, Montpellier (Dr. Crochet), and Department of Gynecological Surgery, Clermont-Ferrand University Hospital Estaing, Clermont-Ferrand (Dr. Rabischong), France*

**ABSTRACT** **Study Objective:** The technical conduct of total laparoscopic hysterectomy (LH) is critical to surgical outcomes. This study explored the validity evidence of an objective scale specific to the assessment of technical skills (H-OSATS) for 7 tasks of an LH with salpingo-oophorectomy procedure performed in the operating room.

**Design:** Observational cohort study.

**Setting:** Two academic hospitals in Marseille and Montpellier, France.

**Patients:** Three groups of operators (novice, intermediate, and experienced surgeons) were video recorded during their live performances of LH on a simple case. For each group, a dozen unedited videos were obtained for the following tasks: division of the round ligament, division of the infundibulopelvic ligament, creation of the bladder flap, opening of the posterior peritoneum, division of the uterine vessels, colpotomy, and closure of the vault.

**Interventions:** Two qualified raters blindly assessed each video using the H-OSATS rating scale. Inter-rater reliability and test-retest reliability were calculated as measures of internal structure. In a separate round of evaluations, the raters provided a global competent/noncompetent decision for each performance. As a measure of consequential validity, a pass/fail score was set for each task using the contrasting group method.

**Measurements and Main Results:** Three tasks (creation of the bladder flap, colpotomy, and closure of the vault) displayed sound validity evidence: a meaningful total score difference among the 3 groups of experience as well as between the intermediate and experienced surgeons, reliability outcomes of  $>0.7$ , and a pass/fail score with a theoretical false-positive rate of  $<10\%$ .

**Conclusion:** The validity evidence of the H-OSATS rating scale differed for separate evaluations of the 7 tasks. Three tasks (i.e., creation of the bladder flap, colpotomy, and closure of the vault) revealed sound validity evidence, including at the level of the attending surgeon, whereas other tasks were more consistent with low-stakes formative evaluation standards.

**Keywords:** Assessment tool; Laparoscopy; Hysterectomy; Procedural skills; Validity

A growing concern in the surgical community is to accurately monitor the technical progress of trainees and

ultimately provide them with a certification for independent practice [1]. Total laparoscopic hysterectomy (LH) is a

The authors declare that they have no conflict of interest.

This study was funded by the Assistance Publique-Hôpitaux de Marseille.

This study was approved by the institutional review board of the French College of Obstetricians and Gynecologists (CEROG 2015-GYN-0801; August 31, 2015).

**Data availability statement:** The authors confirm that the data supporting the findings of this study are available within the article and its appendices.

Corresponding author: Patrice Crochet, PhD, MD, Department of Obstetrics and Gynecology, La Conception Hospital, Aix Marseille University, 147 Blvd Baille, 13005 Marseille, France.

E-mail: [pcrochet.marseille@gmail.com](mailto:pcrochet.marseille@gmail.com)

complex procedure that is mainly taught in the operating room (OR) by mentoring [2,3]. Yet, there is no consensus on how to assess competence for this procedure. Technical performance can be directly assessed in a structured manner using rating scales. Thus, a qualified observer rates an operator on predefined aspects of operative performance. Analysis of video-recorded procedures using generic global rating scales demonstrated a good correlation between a surgeon's skills and early complications, using criteria such as handling of instruments and respect for tissue [4–6]. However, generic global rating scales are not designed to encompass all the domains of intraoperative performance that lead to safety and long-term effective outcomes. Procedure-specific rating scales were developed to provide informative feedback and to provide assessment criteria that could be linked to the safety and effectiveness of a given procedure [7–9]. H-OSATS is a detailed procedure-specific rating scale that was designed to assess operative performances of the LH procedure with or without salpingo-oophorectomy, with a view to its potential use for both formative and summative purposes [10].

The objective of this study was to explore the validity evidence of 7 core tasks of the H-OSATS rating scale in the OR setting.

## Materials and Methods

This observational cohort study was undertaken at the academic gynecology departments of La Conception Hospital (Marseille) and Arnaud de Villeneuve Hospital (Montpellier) in France. The study was conducted between January 2016 and April 2019. Informed oral consent was obtained from the surgeons assessed. Patients were provided informed consent about the video recordings at the preoperative consultation. All the videos were made anonymous. Ethical approval was obtained from the institutional review board of the French College of Obstetricians and Gynecologists (CEROG 2015-GYN-0801).

### *Case Selection and Procedure Standardization*

LH procedures were recorded in the OR using a standard rod lens laparoscope with a detachable camera. The inclusion criteria were set to ensure a homogeneous level of difficulty among the operators: simple cases of patients presenting with a pear-shaped uterus (maximum size: resembling 8 weeks' pregnancy) involving minimal or no pelvic adhesion. All cases were undertaken using a uterine manipulator: Clermont-Ferrand (Karl Storz, Tuttingen, Germany) or RUMI II (CooperSurgical, Trumbull, CT). The type of energy device used was left to the discretion of the surgeon (bipolar electrosurgery or ultrasonic). The operator was always placed on the left side of the patient.

### *List of Tasks Assessed*

Six core tasks of the LH procedure were considered, as defined by the procedure-specific H-OSATS rating scale [10]: (1) division of the round ligaments, (2) creation of the bladder flap, (3) opening of the posterior peritoneum, (4) division of the uterine vessels, (5) colpotomy, and (6) retrieval of the uterus and closure of the vault. A seventh task performed in the case of LH with salpingo-oophorectomy was evaluated: division of the infundibulopelvic (IP) ligament. Three tasks were unilaterally considered (right- or left-sided): (1) division of the round ligament, (2) division of the IP ligament, and (3) opening of the posterior peritoneum. All "division of the uterine vessels" videos were left-sided because we assumed that the evaluation of this task was easier on the left side in the context of a blinded video review. Only 2 out of the 5 items included in the "retrieval of the uterus and closure of the vault" videos were retained for skills assessment on the videos, and consequently this task was relabeled "closure of the vault." To allow a more accurate comparison among the operators, all "closure of the vault" videos concerned cases where the uterosacral ligaments' insertion into the vagina was intact after colpotomy.

### *Participants*

Three groups of surgeons with different levels of experience (novice, intermediate, and experienced surgeons) were video recorded during LH performances. The inclusion criteria for the novice group were defined as follows: post-graduate gynecology residents (from year 3 to 5), attendance at a lecture on a step-by-step description of the LH procedure, open access to a didactic training presentation and a video demonstration of an experienced surgeon performing the LH procedure, prior completion of the fundamentals of laparoscopic surgery curriculum on a fundamentals of laparoscopic surgery box trainer during a 6-month rotation in a gynecology department, and prior experience as a primary operator in 1 to 4 LH procedures [11]. Videos from the intermediate group were obtained from first- and second-year attending surgeons who had performed between 10 and 50 LH procedures independently. Videos from the experienced group were obtained from attending surgeons with more than 5 years of experience in an academic hospital and roles in resident and young attending surgeons' education. They had performed more than 200 LH procedures.

### *Video Selection*

Videos from the novice group were selected using the following criteria: the task was entirely performed by a novice surgeon without any proactive guidance from the attending surgeon. When guidance of the attending surgeon was necessary to ensure patient safety, the degree of this

intervention was evaluated by the attending surgeon at the end of the procedure using the Zwisch scale. This scale defines the level of guidance provided to the resident during a case, and it has 4 levels: (1) show and tell, (2) active help, (3) passive help, and (4) supervision only [12]. All videos requesting more than passive help were excluded. Consequently, videos from the novice group selected for evaluation could be extracted using video procedures where the operator had performed only certain chosen tasks. Surgeons from the intermediate and experienced groups received no guidance during the procedures.

Eligible videos of the LH procedures were prospectively collected. One of the authors (P.C.) screened each video to determine which tasks were appropriate for video review (i. e., tasks recorded from the beginning to the end with no cut). These videos were subsequently edited to obtain the full sequence of eligible tasks. Data collection ended when 9 to 11 videos of each task were obtained for each of the 3 groups with different levels of experience. Among the eligible videos of tasks, there was a video of an organ injury that occurred (e.g., bladder) during a procedure performed by an operator from the intermediate group. This video was excluded from the analysis because of a potential bias in favor of the novice group that operated under guidance for safety reasons.

### **Assessment**

All the videos were assessed independently by 2 raters using the H-OSATS rating scale. These raters were attending surgeons at a teaching hospital. They had completed more than 50 LHs and had experience in teaching LH at their institution to trainees as well as in society-related courses. Each item of a specific task was rated from 1 to 5 (with 1 indicating an item not performed or dangerously performed, 3 indicating an item performed adequately with no obvious danger for the patient, and 5 indicating an ideal performance). The raters assessed the videos and applied these rating criteria to each item independently of each other. A calibration session was held before the video-based assessment to ensure that the interpretation of performances as defined by the scale was similar for the 2 raters. Any identifiers were removed from the laparoscopic videos, and the raters were blinded to the identity of the operator. The order of the videos to be reviewed was changed for each round of ratings.

### **Validity Evidence**

Validity evidence for the evaluation of each of the 7 tasks was assessed in accordance with Messick's conceptual framework of validity [13,14]. Evidence of content had been established by agreement among a Delphi panel of 14 international experts, which was published by our team in 2018 [10]. The response process was ensured by the calibration session between the 2 raters. We examined the validity

evidence of 3 other main sources of validity: (1) relationship to other variables, (2) internal structure, and (3) consequences. Evidence to support the relationship to other variables was explored by comparing the scores of the 3 groups having different levels of experience. The proof of internal structure was provided by evaluating the reliability of the H-OSATS rating scale: the inter-rater reliability was assessed by correlating scores between the 2 independent raters. The test–retest reliability was evaluated by correlating the scores assessed at 2 different time points—at least 1 month apart—by 1 of the raters. The raters were asked not to spend more than 6 hours per day performing the video reviews to avoid evaluation discrepancies due to boredom. Evidence of consequences was based on a pass/fail competency standard, which was established using the contrasting group method [15,16]. The performance of the participants was first scored by the rater using the H-OSATS rating scale, as explained previously. One month later, the same rater assessed the overall performance of the task, providing a competent/noncompetent classification. In the context of the present study, competence was defined as the ability to safely and effectively perform a task. In case of a disagreement between the 2 raters, a third rater assessed the performance to assign it to the competent/noncompetent group. Next, the operators were divided into 2 contrasting groups (i.e., competent and noncompetent). When using the contrasting group method, the cutoff point was set by identifying the intercept of the normally distributed curves that represented the task-score distributions of the groups defined by their competent/noncompetent status. Theoretical false-positive and false-negative figures were reported because of the small group sizes. Indeed, observed false-positive and false-negative rates are very sensitive to small samples and outliers [15].

### **Statistics**

The results were reported as their mean values and standard derivation. The sample size was calculated using an analysis of variance in PASS 2008 software v.08.0.7 (NCSS, LLC, Kaysville, UT), with the hypothesis of a mean OSATS score of 25 (out of 30) in the experienced group, 20 in the intermediate group, and 13 in the novice group (80% statistical power, type I error rate of 5%). The standard deviation of mean values was assumed to be 4.92, and the common standard deviation within a group was assumed to be 8.00. These values were chosen on the basis of the results of a prior study on LH [17]. Using these criteria, 10 participants had to be recruited in each group. To compare scores among the groups and control for the variation coming from the centers, a linear mixed model was used. The inter-rater reliability was assessed by correlating the total and item scores between the 2 independent raters, using the intraclass correlation coefficient (ICC) (2-way mixed-effects model, absolute agreement). The test–retest reliability was assessed by correlating the scores

assessed at 2 different time points by the same rater, using the ICC. Data were analyzed using SPSS v.20.0 (IBM Corp., Armonk, New York), and  $p < .05$  was considered statistically significant.

## Results

### Operative Time

For each task, the mean operative times of the video sequences and comparison of the 3 groups with different levels of experience are presented in [Table 1](#).

### Relation to Other Variables

The task scores and comparison of the 3 groups are provided in [Table 2](#). The total scores were different for all tasks, except “division of the IP ligament.” Each individual item score was significantly different for the tasks “creation of the bladder flap,” “opening of the posterior peritoneum,” “colpotomy,” and “closure of the vault.” None of the items regarding lateral uterine manipulation found a difference among the 3 groups, except for “creation of the bladder flap.”

A group-by-group comparison is provided in [Supplemental Table 1](#) for all tasks. The total scores were higher in the experienced group than in the intermediate group for “creation of the bladder flap,” “opening of the posterior peritoneum,” “colpotomy,” and “closure of the vault.” All item scores were higher in the experienced group than in the intermediate group for “opening of the posterior peritoneum,” “colpotomy,” and “closure of the vault.” None of the item scores differed between the experienced and intermediate groups for “division of the IP ligament.” The total scores were higher in the intermediate group than in the novice group for “creation of the bladder flap,” “division of the uterine vessels,” and “closure of the vault.” All item scores were higher in the intermediate group than in the novice group for “closure of the vault.” None of the item scores for “division of the IP ligament,” “opening of the posterior peritoneum,” and “colpotomy”

were different in the intermediate group compared with the novice group.

Sixty-four percent (139/217) of the videos were recorded at the first center (Marseille) and 36% (78/217) at the second center (Montpellier). The estimation of variance due to the center effect was 19% for the “division of the round ligament” task and 0.1% for the “closure of the vault” task, and there was no variance for the 5 other tasks.

### Internal Structure

The inter-rater reliability ICCs for total scores were  $>0.7$  for all tasks, except “opening of the posterior peritoneum” ([Table 3](#)). These coefficients for individual items of a specific task ranged from 0.389 ( $p = .090$ ) to 0.861 ( $p \leq .001$ ). Five out of the 7 items related to uterine manipulation or exposure had an inter-rater reliability ICC of  $<0.6$ . The test-retest reliability coefficients for total scores were  $>0.7$  for all tasks, except “opening of the posterior peritoneum.” These coefficients for individual items of a specific task ranged from 0.437 ( $p = .049$ ) to 0.962 ( $p < .001$ ).

### Consequences

The cutoff scores obtained by the contrasting group method for each of the 7 tasks are displayed in [Fig. 1](#). Theoretical false-positive and false-negative rates are presented in [Table 4](#). “Creation of the bladder flap,” “colpotomy,” and “closure of the vault” had a theoretical false-positive rate of  $<10\%$ . “Colpotomy” had a theoretical false-negative rate of  $<10\%$ . “Division of the IP ligament” and “opening of the posterior peritoneum” had a theoretical false-positive or false-negative rate of  $>25\%$ . Details about competence status (sorted by groups with different levels of experience) and observed false-negative and false-positive rates can be found in [Supplement Table 2](#).

## Discussion

This study gathered the validity evidence of an assessment tool specific to LH with salpingo-oophorectomy in the

**Table 1**

Operative time (in minutes) of the 7 core tasks of the laparoscopic hysterectomy procedure. Results are presented as the mean (standard deviation). Comparison of the 3 groups of operators by a linear mixed model

H-OSATS tasks	Novice	Intermediate	Experienced	p-value
Division of the round ligament	4.8 (2.4)	2.4 (1.2)	1.7 (0.9)	$<.001$
Division of IP ligament	8.6 (8.4)	5.1 (2.7)	3.7 (2.2)	.112
Creation of the bladder flap	19.6 (7.1)	10.9 (5.7)	6.7 (4.6)	$<.001$
Opening of the posterior peritoneum	3.2 (2.2)	2.7 (1.2)	2 (1)	.210
Division of the left uterine vessels	14.3 (9)	6.6 (3.6)	5 (1.9)	.001
Colpotomy	21.5 (7.8)	17.2 (7.4)	9 (4.2)	.001
Closure of the vault	32.5 (8)	22.5 (7)	15.4 (5.3)	$<.001$

H-OSATS = objective scale specific for the assessment of technical skills for laparoscopic hysterectomy; IP = infundibulopelvic.

**Table 2**

Scores obtained by the 3 groups of operators for the 7 tasks of laparoscopic hysterectomy with salpingo-oophorectomy and corresponding items. Results are presented as mean (standard derivation). Comparison by a linear mixed model

**Division of the round ligament**

	Novice N = 10	Intermediate N = 9	Experienced N = 11	p-value
Manipulator: push uterus cranially and laterally toward the opposite side (out of 5)	3.7 (1.3)	3.3 (1.1)	3.7 (1.0)	.861
Coagulation and transection of the round ligament (out of 5)	3.5 (1.1)	4.1 (0.5)	4.6 (0.5)	<.001
Individualize the anterior and posterior leaf of the broad ligament (out of 5)	3.3 (0.6)	3.8 (0.6)	4.1 (0.5)	<.001
Total (out of 15)	10.4 (2.7)	11.2 (1.9)	12.4 (1.8)	.022

**Division of the IP ligament**

	Novice N = 10	Intermediate N = 11	Experienced N = 10	p-value
Manipulator: push uterus cranially and laterally toward the opposite side (out of 5)	2.9 (1.1)	2.8 (0.9)	3.3 (0.9)	.423
Expose IP ligament (out of 5)	3.5 (1.0)	3.6 (0.6)	3.9 (0.9)	.531
Ureter identification (out of 5)*	2.7 (1.4)	3.4 (1.4)	3.8 (1.1)	.042
Coagulate using an appropriate energy source or suture (out of 5)	3.5 (1.0)	3.7 (0.9)	3.8 (1.0)	.708
Section the IP ligament (out of 5)	3.5 (0.9)	3.6 (0.6)	4.1 (0.9)	.082
Total (out of 25)	16.0 (4.5)	17.1 (3.0)	18.6 (3.7)	.155

**Creation of the bladder flap**

	Novice N = 11	Intermediate N = 9	Experienced N = 10	p-value
Manipulator: push uterus cranially and laterally to the opposite side (left and right) (out of 10)	6.4 (1.4)	6.6 (1.4)	7.7 (1.3)	.046
Open the anterior leaf of the broad ligament on both sides down to the level of the vesicouterine reflection (out of 10)	5.3 (1.4)	7.4 (1.0)	7.9 (1.3)	<.001
Manipulator: push uterus cranially and posteriorly (out of 5)	2.8 (0.7)	3.4 (0.7)	4.3 (0.5)	<.001
Section the peritoneum in front of the lower uterine segment (out of 5)	2.6 (0.6)	3.6 (0.9)	4.2 (0.6)	<.001
The bladder is grasped at the midline, applying an anterior-superior traction (out of 5)	2.5 (0.5)	3.2 (0.5)	4.2 (0.5)	<.001
Opening of the vesicouterine space at the midline until the cervicovaginal margin is exposed (out of 5)	2.5 (0.7)	3.6 (0.9)	4.2 (0.6)	<.001
Total (out of 40)	22.4 (3.9)	27.7 (4.2)	32.3 (3.4)	<.001

**Opening of the posterior peritoneum**

	Novice N = 10	Intermediate N = 10	Experienced N = 11	p-value
Manipulator: push uterus anteriorly and cranially toward the opposite side (out of 5)	3.5 (1.0)	3.4 (0.9)	4.1 (0.5)	.002
Dissection and section of the posterior leaf of the broad ligament toward the insertion of the uterosacral ligament (out of 5)	2.7 (0.8)	3.0 (0.5)	4.0 (0.6)	<.001
Total (out of 10)	6.2 (1.3)	6.4 (1.2)	8.1 (0.8)	<.001

**Division of the uterine vessels**

	Novice N = 11	Intermediate N = 10	Experienced N = 10	p-value
Manipulator: push uterus cranially and to the opposite side (out of 5)	3.9 (0.5)	3.8 (0.8)	4.3 (0.7)	.118
Optimize exposure of the uterine vessels (out of 5)	3.9 (0.6)	3.9 (0.9)	4.3 (0.4)	.199
Skeletonize uterine vessels at the ascending portion of the uterine artery (out of 5)	2.5 (0.8)	3.4 (0.9)	3.8 (0.8)	<.001
Coagulate the uterine vessels using an appropriate energy source or suture (out of 5)	2.8 (0.9)	3.6 (1.1)	3.6 (0.9)	.038
Section uterine vessels in the ascending portion, at the level of the colpotomizer (out of 5)	2.8 (0.8)	3.6 (0.6)	3.6 (0.8)	.003
Divide cervical attachments of the cardinal ligaments (out of 5)	2.7 (0.7)	3.0 (0.6)	3.9 (0.7)	<.001
Total (out of 30)	18.5 (3.5)	21.1 (4.3)	23.3 (3.3)	.001

<b>Colpotomy</b>				
	Novice N = 12	Intermediate N = 10	Experienced N = 10	p-value
Manipulator: push uterus cranially (out of 5)	3.5 (0.5)	3.4 (0.7)	4.5 (0.5)	<.001
Identify the cervicovaginal delineation from the colpotomizer (out of 5)	3.0 (0.6)	3.0 (0.9)	4.4 (0.6)	<.001
Check that there are no interposed elements around the vaginal fornices, and complete dissection if necessary (out of 5)	2.8 (0.8)	3.0 (0.9)	4.3 (0.6)	<.001
Proceed to circumferential colpotomy using an appropriate energy source (out of 5)	2.5 (0.5)	3.0 (0.9)	4.1 (0.6)	<.001
Total (out of 20)	11.9 (1.9)	12.4 (3.2)	17.1 (1.5)	<.001
<b>Closure of the vault</b>				
	Novice N = 11	Intermediate N = 10	Experienced N = 11	p-value
Suture the vaginal vault with interrupted or continuous sutures (out of 5)	1.9 (0.4)	2.8 (0.8)	4.1 (0.5)	<.001
Vaginal suture including sufficient width of vaginal mucosa and fascia (out of 5)	2.0 (0.5)	2.8 (0.7)	4.0 (0.7)	<.001
Total (out of 10)	3.9 (0.7)	5.6 (1.2)	8.1 (1.0)	<.001
H-OSATS = objective scale specific for the assessment of technical skills for laparoscopic hysterectomy; IP = infundibulopelvic. * With or without fenestration as described in the H-OSATS scale.				

OR. Separate evaluations of the core tasks of the procedure allowed exploration of the assessment potential of each task independently, with minimal interference from what was performed before. Validity evidence differed across the tasks. Three tasks (i.e., creation of the bladder flap, colpotomy, and closure of the vault) emerged by their good results across the 3 tested aspects of validity: (1) a meaningful total score difference among the 3 groups with different levels of experience and between the 2 groups of attending surgeons, (2) reliability outcomes of  $>0.7$ , and (3) a pass/fail score obtained with a theoretical false-positive rate of  $<10\%$ . These 3 tasks have common characteristics that may explain these outcomes: first, the operative time was higher than for the other tasks, which gives the rater more comfort in making a judgment. Second, a good understanding of anatomic landmarks is crucial for these tasks, and any error or approximation is easily spotted by the rater. Finally, these tasks require a high level of generic skills. Surprisingly, “division of the IP ligament” could not be discriminated among the 3 groups with different levels of experience. This can be explained not only because different techniques were allowed (with or without fenestration), but also because it was not easy to ascertain from a mute video-based assessment whether the surrounding structures were identified by the operator. The “division of the utero-ovarian ligament” task, performed in cases of hysterectomy without salpingo-oophorectomy, was not considered in this study and needs further investigation. The overall compelling findings of reliability, using only 2 raters, showed that the tool is suitable for providing formative feedback

because a reliability coefficient of  $>0.70$  is deemed acceptable [18]. The inter-rater reliability may be improved by increasing the number of raters or rated procedures per surgeon [19].

Tremblay et al [20] published a comprehensive rating scale specific to the LH procedure, but no validity evidence was provided in the surgical setting. Recently, Savran et al [21] published a short version of a procedure-specific scale. This scale revealed sound validity evidence when used for a full procedure in the OR with 2 groups of operators: beginners and experienced surgeons. Application of this scale for summative purposes needs further evidence because the beginners scored quite low, and it was unclear which level of guidance was received by this group.

An ideal evaluation of technical skills for LH should include all tasks of a full-length procedure. In this case, proficiency would be achieved only if all tasks were individually rated above the cutoff scores. However, certain tasks (i.e., creation of the bladder flap, colpotomy, and closure of the vault) demonstrated higher validity evidence than others (e.g., division of the IP ligament). A peer review limited to these 3 specific tasks could be advantageous to focus the rater’s attention on the more discriminative part of the procedure. This hypothesis should be confirmed by further studies. In addition, the postoperative consequences of the ratings should be further investigated because the cutoff scores were based on an intraoperative evaluation of competence. Although performances below the cutoff score can potentially be predictive of increased adverse events or costs, this study did not provide evidence on this matter.

**Table 3**

Reliability coefficients for total scores of the 7 core tasks of laparoscopic hysterectomy with salpingo-oophorectomy

	Inter-rater	Test-retest
Division of the round ligament	0.832 (p <.001)	0.839 (p <.001)
Division of the IP ligament	0.703 (p = .001)	0.842 (p <.001)
Creation of the bladder flap	0.813 (p <.001)	0.860 (p <.001)
Opening of the posterior peritoneum	0.512 (p = .021)	0.669 (p <.001)
Division of the uterine vessels	0.743 (p <.001)	0.967 (p <.001)
Colpotomy	0.759 (p <.001)	0.775 (p <.001)
Closure of the vault	0.858 (p <.001)	0.886 (p <.001)

IP = infundibulopelvic.

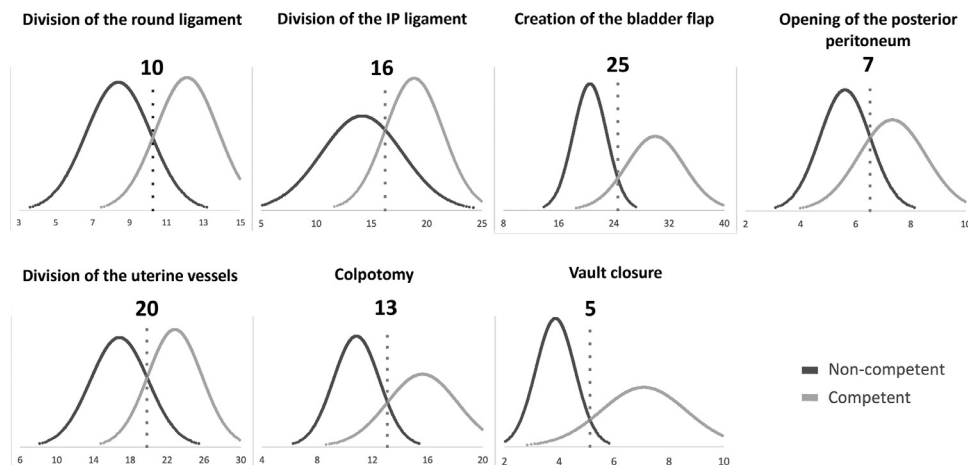
The H-OSATS scale was revealed to be a valuable formative assessment tool in the OR for simple cases. In their practice, surgeons are often confronted by a wider spectrum of anatomic presentations that were not tested in the present study. Consequently, this scale cannot be recommended as the sole certification tool for the LH procedure. Its role in a future multimodal process of summative assessment needs to be further investigated.

The strength of the present study lies in the incremental levels of experience of the 3 groups of operators. The participation of the novice group with minimal guidance was ethically possible only because the scale was evaluated task by task. Careful selection of video-recorded task performances in the novice group could not fully eliminate the guidance bias that remained for safety reasons. Another limitation was the absence of information

regarding the handedness of the operators, given that all operators stood on the left side of the patient. Assessment through a video interface limits the scope of evaluation of operative skills [22]. Thus, some facets of technical competence, such as setting exposure, can be difficult to measure by watching an internal video feed alone [9]. This was confirmed in this study by the poor performance of items related to exposure or uterine manipulation. Finally, the broad range of experience permitted in the intermediate group (10 to 50 LHs) implied that some participants had not completed their learning curve. Thus, some participants did not reach the cutoff score for competence in this group. However, the scale found differences in performance between the 2 groups of attending surgeons (i.e., <2 years or >5 years of experience) for most of the tasks.

**Fig. 1**

Contrasting group curves for each of the 7 core tasks of laparoscopic hysterectomy with salpingo-oophorectomy. The pass/fail scores are represented by the dashed line.





**Table 4**

Theoretical false-positive and false-negative rates for the 7 core tasks of the laparoscopic hysterectomy procedure, calculated using the normally distributed curves that represent the score distributions of the competent and noncompetent groups

	False-positive rate, %	False-negative rate, %
Division of the round ligament	14.3	13.6
Division of the IP ligament	28.4	15.9
Creation of the bladder flap	4.8	9.7
Opening of the posterior peritoneum	15.8	25.5
Division of the uterine vessels	16.9	15.1
Colpotomy	8.7	16
Closure of the vault	3.8	10.3

IP = infundibulopelvic.

## Conclusion

The H-OSATS rating scale proved to be a promising video-based assessment tool in the OR for 3 core parts of the procedure: (1) creation of the bladder flap, (2) colpotomy, and (3) closure of the vault. Other parts of the H-OSATS rating scale showed mixed outcomes and could probably not be considered individually for high-stakes assessment. The next steps would involve confirming the value of H-OSATS individual tasks in a multicenter setting and exploring the assessment potential of this scale during full-length procedures in the OR.

## References

- Feldman LS, Pryor AD, Gardner AK, et al. SAGES video-based assessment (VBA) program: a vision for life-long learning for surgeons. *Surg Endosc.* 2020;34:3285–3288.
- Desai VB, Xu X. An update on inpatient hysterectomy routes in the United States. *Am J Obstet Gynecol.* 2015;213:742–743.
- Igwe E, Hernandez E, Rose S, Uppal S. Resident participation in laparoscopic hysterectomy: impact of trainee involvement on operative times and surgical outcomes. *Am J Obstet Gynecol.* 2014;211:484.e1–484.e7.
- Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997;84:273–278.
- Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg.* 2005;190:107–113.
- Vaidya A, Aydin A, Ridgley J, Raison N, Dasgupta P, Ahmed K. Current status of technical skills assessment tools in surgery: a systematic review. *J Surg Res.* 2020;246:342–378.
- Zevin B, Bonrath EM, Aggarwal R, et al. Development, feasibility, validity, and reliability of a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery. *J Am Coll Surg.* 2013;216:955–1033.
- Bonrath EM, Dedy NJ, Gordon LE, Grantcharov TP. Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. *Ann Surg.* 2015;262:205–212.
- Champagne BJ, Steele SR, Hendren SK, et al. The American Society of Colon and Rectal Surgeons assessment tool for performance of laparoscopic colectomy. *Dis Colon Rectum.* 2017;60:738–744.
- Knight S, Aggarwal R, Agostini A, Loundou A, Berdah S, Crochet P. Development of an objective assessment tool for total laparoscopic hysterectomy: a Delphi method among experts and evaluation on a virtual reality simulator. *PLoS One.* 2018;13:e0190580.
- Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial. *Am J Surg.* 2010;199:115–120.
- George BC, Teitelbaum EN, Meyerson SL, et al. Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. *J Surg Educ.* 2014;71:e90–e96.
- Association AER, association AP, NCME. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.
- Ghaderi I, Manji F, Park YS, et al. Technical skills assessment toolbox: a review using the unitary framework of validity. *Ann Surg.* 2015;261:251–262.
- Jørgensen M, Konge L, Subhi Y. Contrasting groups' standard setting for consequences analysis in validity studies: reporting considerations. *Adv Simul (Lond).* 2018;3:5.
- Szasz P, Bonrath EM, Louridas M, et al. Setting performance standards for technical and nontechnical competence in general surgery. *Ann Surg.* 2017;266:1–7.
- Crochet P, Aggarwal R, Knight S, Berdah S, Boubli L, Agostini A. Development of an evidence-based training program for laparoscopic hysterectomy on a virtual reality simulator. *Surg Endosc.* 2017;31:2474–2482.
- Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ.* 2004;38:1006–1012.
- Konge L, Annema J, Clementsen P, Minddal V, Vilmann P, Ringsted C. Using virtual-reality simulation to assess performance in endobronchial ultrasound. *Respiration.* 2013;86:59–65.
- Tremblay C, Grantcharov T, Urquia ML, Satkunaratnam A. Assessment tool for total laparoscopic hysterectomy: a Delphi consensus survey among international experts. *J Obstet Gynaecol Can.* 2014;36:1014–1023.
- Savran MM, Hoffmann E, Konge L, Ottosen C, Larsen CR. Objective assessment of total laparoscopic hysterectomy: development and validation of a feasible rating scale for formative and summative feedback. *Eur J Obstet Gynecol Reprod Biol.* 2019;237:74–78.
- Madani A, Vassiliou MC, Watanabe Y, et al. What are the principles that guide behaviors in the operating room?: creating a framework to define and measure performance. *Ann Surg.* 2017;265:255–267.

## Supplemental Table 1

Group by group comparison of performances (mean + standard derivation) between novice, intermediate and experienced groups (Comparison by a linear mixed model)

Tasks	items	Novices	Intermediates	Experienced	Novices	Novices	Intermediates
					vs	vs	vs
					Intermediates	Experienced	Experienced
<b>Division of the round ligament</b>	Manipulator: push uterus cranially and laterally towards the opposite side (5)	3.7 (1.3)	3.3 (1.1)	3.7 (1.0)	.610	.945	.659
	Coagulation and transection of the round ligament (5)	3.5 (1.1)	4.1 (0.5)	4.6 (0.5)	.013	<.001	.163
	Individualize the anterior and posterior leaf of the broad ligament (5)	3.3 (0.6)	3.8 (0.6)	4.1 (0.5)	.020	<.001	.086
	Total (15)	10.4 (2.7)	11.2 (1.9)	12.4 (1.8)	.144	.006	.228
<b>Division of IP ligament</b>	Manipulator: push uterus cranially and laterally towards the opposite side (5)	2.9 (1.1)	2.8 (0.9)	3.3 (0.9)	.726	.365	.203
	Expose IP ligament (5)	3.5 (1.0)	3.6 (0.6)	3.9 (0.9)	.758	.278	.418
	Ureter identification (5)	2.7 (1.4)	3.4 (1.4)	3.8 (1.1)	.092	.014	.381
	Coagulate using an appropriate energy source or suture (5)	3.5 (1.0)	3.7 (0.9)	3.8 (1.0)	.610	.412	.740
	Section the IP ligament (5)	3.5 (0.9)	3.6 (0.6)	4.1 (0.9)	.518	.031	.111
	Total (25)	16.0 (4.5)	17.1 (3.0)	18.6 (3.7)	.405	.055	.250
<b>Creation of the bladder flap</b>	Manipulator: push uterus cranially and laterally to the opposite side (left and right) (10)	6.4 (1.4)	6.6 (1.4)	7.7 (1.3)	.787	.021	.052
	Open the anterior leaf of the broad ligament on both sides down to the level of the vesico-uterine reflection (10)	5.3 (1.4)	7.4 (1.0)	7.9 (1.3)	<.001	<.001	.274
	Manipulator: push uterus cranially and posteriorly (5)	2.8 (0.7)	3.4 (0.7)	4.3 (0.5)	.012	<.001	.004
	Section the anterior peritoneum in front of the lower uterine segment (5)	2.6 (0.6)	3.6 (0.9)	4.2 (0.6)	.001	<.001	.022
	The bladder is grasped at the midline, applying an anterior-superior traction (5)	2.5 (0.5)	3.2 (0.5)	4.2 (0.5)	.003	<.001	<.001
	Opening of the vesico-uterine space at the midline until the cervico-vaginal margin is exposed (5)	2.5 (0.7)	3.6 (0.9)	4.2 (0.6)	<.001	<.001	.018
	Total (40)	22.4 (3.9)	27.7 (4.2)	32.3 (3.4)	<.001	<.001	.002
	Manipulator: push uterus anteriorly and cranially towards the opposite side (5)	3.5 (1.0)	3.4 (0.9)	4.1 (0.5)	.351	.010	.001
<b>Opening of the posterior peritoneum</b>	Dissection and section of the posterior leaf of the broad ligament towards the insertion of the utero-sacral ligament (5)	2.7 (0.8)	3.0 (0.5)	4.0 (0.6)	.088	>.001	.006
	Total (10)	6.2 (1.3)	6.4 (1.2)	8.1 (0.8)	.633	<.001	<.001
<b>Division of the uterine vessels</b>	Manipulator: push uterus cranially and to the opposite side	3.9 (0.5)	3.8 (0.8)	4.3 (0.7)	.648	.114	.052
	Optimize exposure of the uterine vessels	3.9 (0.6)	3.9 (0.9)	4.3 (0.4)	.800	.148	.098
	Skeletonize uterine vessels at the ascending portion of the uterine artery	2.5 (0.8)	3.4 (0.9)	3.8 (0.8)	.003	<.001	.190
	Coagulate the uterine vessels using an appropriate energy source or suture	2.8 (0.9)	3.6 (1.1)	3.6 (0.9)	.028	.028	1.0
	Section uterine vessels in the ascending portion, at the level of the colpotomizer	2.8 (0.8)	3.6 (0.6)	3.6 (0.8)	.008	.002	.555
	Divide cervical attachments of the cardinal ligaments	2.7 (0.7)	3.0 (0.6)	3.9 (0.7)	.210	<.001	.002
	Total (30)	18.5 (3.5)	21.1 (4.3)	23.3 (3.3)	.039	<.001	.087
<b>Colpotomy</b>	Manipulator: push uterus cranially	3.5 (0.5)	3.4 (0.7)	4.5 (0.5)	.531	<.001	<.001
	Identify the cervico-vaginal delineation from the colpotomizer	3.0 (0.6)	3.0 (0.9)	4.4 (0.6)	.974	<.001	<.001
	Check that there are no interposed elements around the vaginal fornices and complete dissection if necessary	2.8 (0.8)	3.0 (0.9)	4.3 (0.6)	.558	<.001	<.001

## Supplemental Table 1

Continued		Novices	Intermediates	Experienced	Novices vs Intermediates	Novices vs Experienced	Intermediates vs Experienced
Tasks	items						
	Proceed to circumferential colpotomy using an appropriate energy source	2.5 (0.5)	3.0 (0.9)	4.1 (0.6)	.069	<.001	<.001
	Total (/20)	11.9 (1.9)	12.4 (3.2)	17.1 (1.5)	.505	<.001	<.001
<b>Closure of the vault</b>	Suture the vaginal vault with interrupted or continuous sutures (/5)	1.9 (0.4)	2.8 (0.8)	4.1 (0.5)	<.001	<.001	<.001
	Vaginal suture including sufficient width of vaginal mucosa and fascia (/5)	2.0 (0.5)	2.8 (0.7)	4.0 (0.7)	0.01	<.001	<.001
	Total (/10)	3.9 (0.7)	5.6 (1.2)	8.1 (1.0)	<.001	<.001	<.001

## Supplemental Table 2

Detail of competence status in the 3 groups of experience for the 7 core tasks of laparoscopic hysterectomy, and observed false-negative and -positive rates

### 1. Division of the round ligament

	Novices	Intermediates	Experienced	Total
Observed false positive rate:	16% (1/6)	Standard of competence: $\geq 10$ points (/15) (1 novice)		
Observed false negative rate:	12.5% (3/24)	(1 experienced, 2 intermediates)		
Competent	5	8	11	24
Non-Competent	5	1	0	6

### 2. Division of IP ligament

	Novices	Intermediates	Experienced	Total
Observed false positive rate:	36% (4/11)	Standard of competence: $\geq 16$ points (/25) (3 novices, 1 intermediate)		
Observed false negative rate:	15% (3/20)	(1 novice, 1 intermediate, 1 experienced)		
Competent	4	8	8	0
Non-Competent	6	3	2	11

### 3. Creation of the bladder flap

	Novices	Intermediates	Experienced	Total
Observed false positive rate:	0% (0/9)	Standard of competence: $\geq 25$ points (/40)		
Observed false negative rate:	9% (2/22)	(1 novice, 1 intermediate)		
Competent	3	8	10	2
Non-Competent	8	1	0	9

### 4. Opening of the posterior peritoneum

	Novices	Intermediates	Experienced	Total
Observed false positive rate:	12.5% (1/8)	Standard of competence: $\geq 7$ points (/10) (1 experienced)		
Observed false negative rate:	39% (9/23)	(3 novices, 6 intermediates)		
Competent	5	8	9	1
Non-Competent	5	3	1	10

<b>5. Division of the left uterine vessels</b>		Standard of competence: $\geq 20$ points (/30)		
Observed false positive rate:	20% (2/10)	(1 novice, 1 experienced)		
Observed false negative rate:	9.5% (2/21)	(1 intermediate, 1 experienced)		
	Novices	Intermediates	Experienced	Total
Competent	3	9	9	1
Non-Competent	8	1	1	10
<b>6. Colpotomy</b>		Standard of competence: $\geq 13$ points (/20)		
Observed false positive rate:	23% (3/13)	(3 novices)		
Observed false negative rate:	16% (3/19)	(1 novice, 2 intermediates)		
	Novices	Intermediates	Experienced	Total
Competent	3	6	10	19
Non-Competent	9	4	0	3
<b>7. Closure of the vault</b>		Standard of competence: $\geq 5$ points (/10)		
Observed false positive rate:	8.3% (1/12)	(1 novice)		
Observed false negative rate:	5% (1/20)	(1 intermediate)		
	Novices	Intermediates	Experienced	Total
Competent	0	9	1	20
Non-Competent	11	1	0	2