Hernia (2014) 18:543–548
DOI 10.1007/s10029-013-1196-2

ORIGINAL ARTICLE

Reliable and valid assessment of Lichtenstein hernia repair skills

C. G. Carlsen · K. Lindorff-Larsen · P. Funch-Jensen · L. Lund · P. Charles · L. Konge

Received: 15 March 2013 / Accepted: 23 November 2013 / Published online: 3 December 2013

Abstract

Purpose Lichtenstein hernia repair is a common surgical procedure and one of the first procedures performed by a surgical trainee. However, formal assessment tools developed for this procedure are few and sparsely validated. The aim of this study was to determine the reliability and validity of an assessment tool designed to measure surgical skills in Lichtenstein hernia repair.

Methods Key issues were identified through a focus group interview. On this basis, an assessment tool with eight items was designed. Ten surgeons and surgical trainees were video recorded while performing Lichtenstein hernia repair, (four experts, three intermediates, and three novices). The videos were blindly and individually assessed by three raters (surgical consultants) using the assessment tool. Based on these assessments, validity and reliability were explored.

Results The internal consistency of the items was high (Cronbach's alpha = 0.97). The inter-rater reliability was very good with an intra-class correlation coefficient (ICC) = 0.93. Generalizability analysis showed a coefficient above 0.8 even with one rater. The coefficient improved to 0.92 if three raters were used. One-way analysis of variance found a significant difference between the three groups which indicates construct validity, $p < 0.001$.

Conclusions Lichtenstein hernia repair skills can be assessed blindly by a single rater in a reliable and valid fashion with the new procedure-specific assessment tool. We recommend this tool for future assessment of trainees performing Lichtenstein hernia repair to ensure that the objectives of competency-based surgical training are met.

Keywords Lichtenstein hernia repair · Validation · Assessment · Surgical training

C. G. Carlsen (✉) · P. Charles
Centre of Medical Education, Aarhus University, Incuba Science Park, Brendstrupgaardsvej 102, 8200 Aarhus N, Denmark
e-mail: cgc@medu.au.dk

K. Lindorff-Larsen
NordSim, Centre for simulation and skills training, Aalborg University Hospital, Hobrovej 18, 9000 Aalborg, Denmark

P. Funch-Jensen
Clinical Institute, Aarhus Hospital, Aarhus University, Brendstrupgaardsvej 100, 8200 Aarhus N, Denmark

L. Lund
Department of Urology, Odense University Hospital, Sdr. Boulevard 29, 5000 Odense, Denmark

L. Konge
Centre for Clinical Education, University of Copenhagen and the Capital Region of Denmark, Blegdamsvej 9, 2100 Copenhagen, Denmark

Purpose

Objective criteria based on employment period or minimal number of performed procedures does not ensure basic surgical competence, as trainees learn at different paces [1]. Competency-based training has therefore become a widespread paradigm in surgical training [2]. Formal assessment is a key feature of the competency-based training to ensure that training objectives are met. Valid assessment tools are therefore needed in surgical training, especially in common and standardized procedures. Lichtenstein hernia repair is a widely standardized and common surgical procedure encountered by most surgical trainees early in their career [3, 4]. The procedure is thoroughly described by Shulman [5] Amid et al. [6] and used world-wide [7, 8]. In spite of these facts, formal assessment in this particular field seems to be poorly developed.

Several tools are available for assessment of laparoscopic procedures [9, 10], but only few assessment tools are available for open surgical procedures although they represent an important part of training. The objective structured assessment of technical skills (OSATS) described by Faulkner and colleagues [11] is widely used for assessment of open surgical procedures. But the OSATS was developed for bench-station training, and use for more complex procedures in the operating room (OR) has therefore resulted in procedure-specific tool development [12–14]. Several tools are used for assessment of Lichtenstein hernia repair, but few are procedure-specific [10, 14–16] and general assessment tools for open surgery are sparsely validated for Lichtenstein hernia repair. Furthermore, the OSATS and tools derived from the OSATS have been used most commonly for direct observation (the assessor being present in the OR). Direct observation could introduce assessment bias [17, 18] due to personal knowledge, among others. Video recordings of Lichtenstein hernia repair provide the possibility of blinded assessment similar to recording of laparoscopic procedures, and they do not require the presence of an experienced assessor at the time of the procedure. Blinded assessment is a desirable alternative to avoid rater bias in surgical training at registrar level [18, 19]. Video recordings, however, have other limitations regarding the design of the tool, and the existing OSATS-derived tools use the items "use of assistant" and "knowledge of instruments" that are difficult to assess in video recordings. A simple and procedure-specific tool for blinded assessment of Lichtenstein hernia repair is therefore needed to provide supervisors with a specific skills assessment tool in surgical hernia repair in a competency-based training curriculum. The aim of this study was to determine the reliability and validity of a new assessment tool designed to assess skills in Lichtenstein hernia repair.

Research questions:

What was the internal consistency of the assessment tool?

What were the inter-rater reliability and generalizability coefficient of the assessment tool?

How was the validity evidence of the assessment tool?

Methods

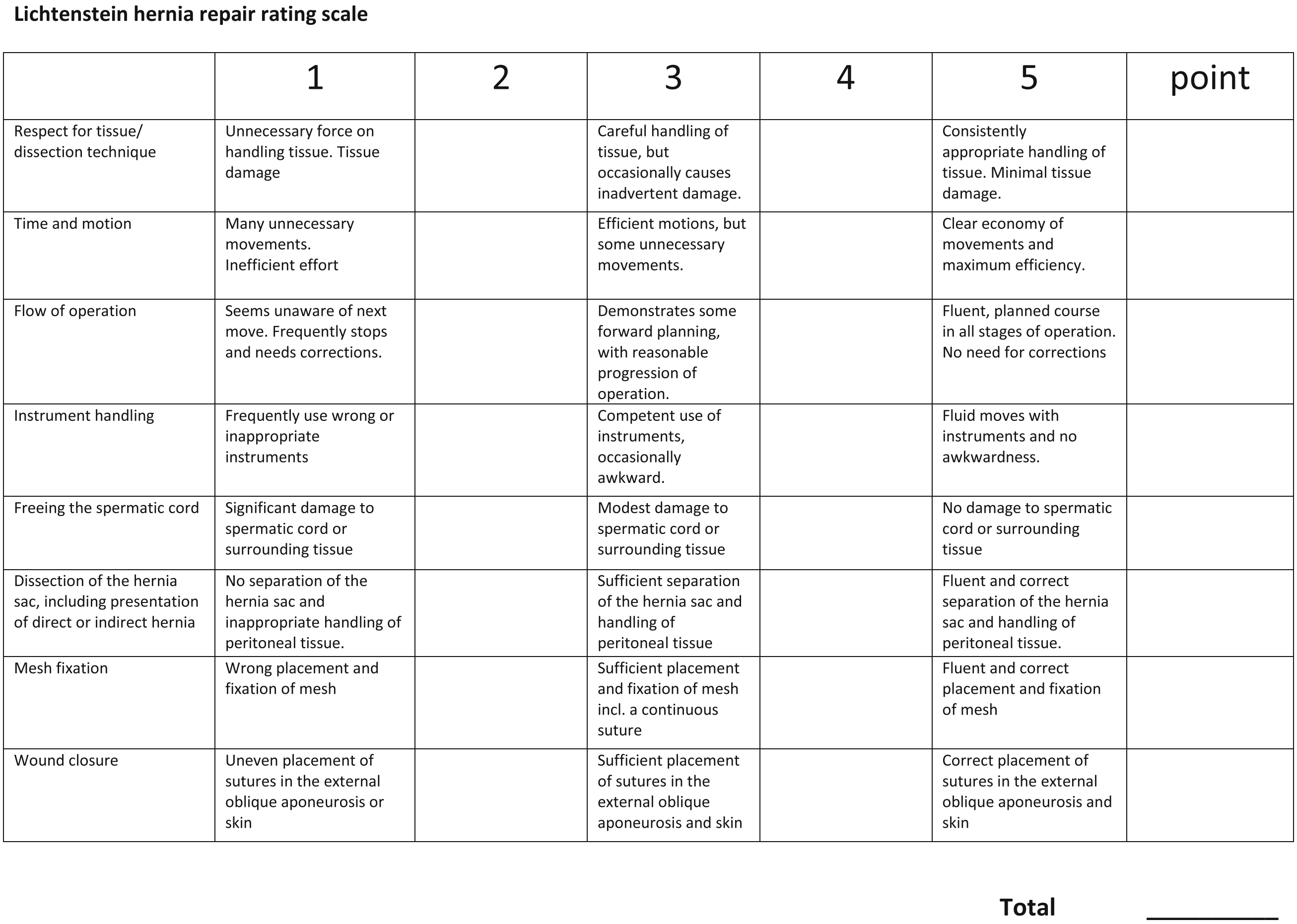
Development of the tool

Based on the OSATS [11, 20], a new tool was designed (Fig. 1). The 5-point Likert scale with anchors described at values 1-3-5 was chosen as the most purposeful scale. The 1-3-5 anchor descriptions follow the tradition in this field [20]. In Denmark the operative description posted by Alex Shulman is used [5] and this bases the tool. A draft of a new tool was made including the OSATS and other existing literature [13–15]. Based on this draft, a focus group interview was conducted. The main purpose of the focus group interview was to identify which key elements of the Lichtenstein hernia repair were to be assessed. The assessment tool used by Larson et al. [14] is more selective in its assessment than most other assessment instruments and devotes particular attention to the surgeon's awareness of the ilioinguinal nerve and femoral vein injury. This issue was widely debated in our group, but the video recordings did not technically allow an adequate assessment of the surgeon's awareness of these anatomical structures. Thus, the nerve may have been identified by the surgeon, but this could not always be ascertained through the video recording. Our group decided to focus on behavior and not on potential injuries, as an otherwise well-performed surgical technique combined with procedural knowledge demonstrates surgical skills, and such skill was believed to protect against major injuries, regardless of kind. Like other tools made for video assessment [12, 13], we left out "knowledge of instruments" and "use of assistant" from the original OSATS. One interview meeting was held comprising the total group and then a new draft tool was made. The group finally identified eight particular skills that should be assessed to fully cover the procedure. Afterward, a team of seven experienced hernia surgeons (FK, KL, LM, MG, HFA, LL, PFJ) commented on the tool. Comments were incorporated until all participants agreed. Finally, a smaller group (KL, LL, PFJ) tested the tool in a pilot study discussing three different video recordings of Lichtenstein hernia repair performed by different surgeons and surgical trainees.

The final tool tested four global skills and four procedure-specific skills. The global skills were: respect for tissue, time and motion, instrument handling and flow of operation. The procedure-specific skills were: dissection of the spermatic cord, presentation of the hernia sac including lateral and medial component, fixation of mesh and wound closure.

Validation process

Ten surgeons and surgical trainees were video recorded performing Lichtenstein hernia repair. They had different levels of experience: four experts (consultants) who had performed >200 Lichtenstein hernia repairs, three intermediates (senior trainees) who had performed 50–75 Lichtenstein hernia repairs, and three novices (junior trainees) who had only little experience (5–10 procedures performed) in Lichtenstein hernia repair. All recordings in



Lichtenstein hernia repair rating scale

	1	2	3	4	5	point
Respect for tissue/ dissection technique	Unnecessary force on handling tissue. Tissue damage		Careful handling of tissue, but occasionally causes inadvertent damage.		Consistently appropriate handling of tissue. Minimal tissue damage.	
Time and motion	Many unnecessary movements. Inefficient effort		Efficient motions, but some unnecessary movements.		Clear economy of movements and maximum efficiency.	
Flow of operation	Seems unaware of next move. Frequently stops and needs corrections.		Demonstrates some forward planning, with reasonable progression of operation.		Fluent, planned course in all stages of operation. No need for corrections	
Instrument handling	Frequently use wrong or inappropriate instruments		Competent use of instruments, occasionally awkward.		Fluid moves with instruments and no awkwardness.	
Freeing the spermatic cord	Significant damage to spermatic cord or surrounding tissue		Modest damage to spermatic cord or surrounding tissue		No damage to spermatic cord or surrounding tissue	
Dissection of the hernia sac, including presentation of direct or indirect hernia	No separation of the hernia sac and inappropriate handling of peritoneal tissue.		Sufficient separation of the hernia sac and handling of peritoneal tissue		Fluent and correct separation of the hernia sac and handling of peritoneal tissue.	
Mesh fixation	Wrong placement and fixation of mesh		Sufficient placement and fixation of mesh incl. a continuous suture		Fluent and correct placement and fixation of mesh	
Wound closure	Uneven placement of sutures in the external oblique aponeurosis or skin		Sufficient placement of sutures in the external oblique aponeurosis and skin		Correct placement of sutures in the external oblique aponeurosis and skin	

Total ___________

Fig. 1 Lichtenstein hernia repair rating scale (translated)

this study concerned males with a primary inguinal hernia diagnosis to ensure a standardized procedure. Patients accepted the recordings prior to the procedure. The video recordings were anonymous, as no patient data were recorded or known to the investigators. Consequently, no permission from the Research Ethics Committee was needed.

All video tapes were recorded in the same manner, recording the operative field only, while standing on the opposite side of the surgeon. The procedure was recorded from skin incision to would closure. Only the visual recordings were assessed; no speech was included. Trainees having performed <10 hernia repairs were supervised by a senior colleague in the OR. The videos were blindly and individually assessed by three raters (surgical consultants) using the newly developed tool. The video recordings only showed the operative field, and it was not possible to identify the age, gender, etc. of the surgeon or of the patient in the recordings. Furthermore, the recordings were randomly numbered with no clue given to experience level of surgeons. Recordings were rated in total, though raters were permitted to use fast forward at their own discretion, as formerly used [21, 22]. Based on these ratings, internal consistency, inter-rater reliability, generalizability analysis and construct validity of the assessment tool were explored.

Statistical analysis

Internal consistency between the items was calculated as Cronbach's alpha. Internal consistency describes whether items measure different parts of the same entity, here hernia repair. Item discrimination index was derived from the average rating across the three rates using item-total correlation (Pearson's correlation). Inter-rater reliability was calculated as an average measures intra-class correlation coefficient (ICC) based on the absolute agreement definition. This describes whether outcome is equal in repeated measures. Generalizability theory was used to describe the optimal balance between the number of raters and the number of items to show both the ability of the tool to discriminate difference in experience and the reproducibility. Therefore, a balanced G-study with procedures (p) crossed with items (i) crossed with raters (r) p × i × r was conducted. The estimated variance components were used to perform a D-study to estimate generalizability coefficients for the number of items and raters used. The aim was a generalizability coefficient above 0.8. Using a mixed-model ANOVA, validity evidence was collected by comparing the raters' average scores in the three defined groups: Novices, intermediates, and experts. The groups were compared two by two using Students *t* test. Statistical analyses were performed using a statistical software package (PASW, version 18.0; SPSS Inc., Chicago, Illinois, USA).

Results

Table 1 shows the raters' mean assessment score in the three groups with different experience when using the new eight-item assessment tool.

The internal consistency was high (Cronbach's alpha = 0.97). The ICC was 0.93. The generalizability analysis showed that three raters achieved a generalizability coefficient of 0.92 with eight items. The D-study estimates a coefficient of 0.81 when a single rater is used (Fig. 2). This is the accepted level.

Table 1 Results of assessment

	Rater 1 mean (SD)	Rater 2 mean (SD)	Rater 3 mean (SD)	Average across raters mean (SD)
Novices	14.7 (2.89)	14.3 (0.58)	15.3 (1.53)	14.8 (1.72)
Intermediates	19.3 (2.08)	17.0 (1.00)	27.0 (5.29)	21.1 (5.37)
Experts	30.8 (3.30)	35.5 (2.52)	36.0 (2.83)	34.1 (3.60)

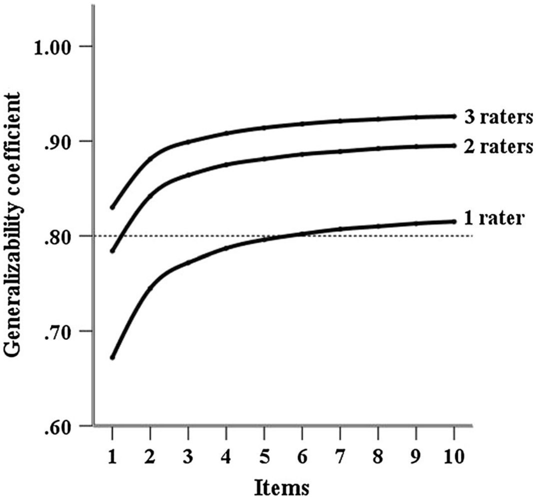


Fig. 2 D-study showing the G-coefficient for different numbers of raters

Table 2 Item discrimination index

Item	Discrimination index
1	0.9604
2	0.9873
3	0.9874
4	0.9824
5	0.9712
6	0.9802
7	0.9259
8	0.9771

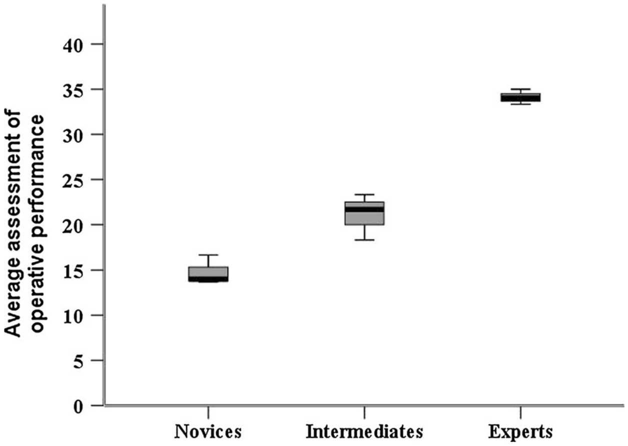


Fig. 3 Average assessment score in the three groups; expert, intermediate and novice

The calculated item discrimination index is shown in Table 2. All eight items discriminate well (above 0.97). This indicates that all items discriminate between novices and experts.

To assess the construct validity of the performers' experience (novices, intermediates, and experts), we performed a mixed-model analysis of variance (ANOVA) using rater average score as a within-group variable and experience as a between-group variable. This test was highly significant ($p < 0.001$).

Furthermore, we conducted an independent samples *t* test comparing experts and intermediates ($p < 0.001$), experts and novices ($p < 0.001$), and novices and intermediates ($p = 0.02$). The results are shown in Fig. 3.

Discussion

The newly developed tool was reliable and valid for assessment of technical skills in Lichtenstein hernia repair. It was able to discriminate between novices', intermediates' and experts' performance. These findings are comparable to similar validations of other OSATS-derived tools [10, 13, 14], though most compare only novices and experts. Our tool was developed for blinded individual rating of video recordings. Other studies found that video tape recordings equaled direct observations [21]. Driscoll et al. [22] found that assessment of video recordings equaled real-time assessments in hernia repair. Video recordings enjoy advantages other than direct observation: for example, raters avoid time-consuming attendance in the OR, the possibility of re-assessment in case of doubt and bias is less likely [18, 19, 23].

A good assessment tool should be reliable, valid and feasible. To obtain feasibility, the number of assessable items should be limited. Our group found that four general items and four procedure-specific items were sufficiently exhaustive to assess video-recorded Lichtenstein hernia repair, as also evidenced by the results. The need for procedure-specific tools versus global scales was discussed by Aggarwal et al. [24]. They concluded that the global score alone provides sufficiently discriminate results. But generic surgical skills must be complemented by procedural knowledge of Lichtenstein hernia repair to ensure that training objectives and patient safety are met as stated by Bell et al. [3]. The experts of the focus groups agreed as to the contents of the tool, which were widely similar to those of other OSATS-derived tools [12–14] adapted to blinded rating. The high internal consistency further supported the selected items. A possible bias could be the Halo effect [25], where the assessor intends to give an opinion as a whole instead of discriminating between separate characteristics. This concerns all global assessment tools and in an attempt to avoid this bias we made three detailed anchor descriptions of each items and used three individual raters.

The ICC of 0.93 was very good and confirmed the reproducibility. The rating was conducted individually. For assessment of video recording, other adaptations of the tool are required, as discussed above, but even with these disadvantages, we preferred the blinded rating for this open surgery procedure for the reasons mentioned above. Video recording is also fully comparable to assessment of laparoscopic procedures, which becomes more and more common in surgical training. The generalizability analysis showed that one rater was sufficient to reliably assess a blinded video. This is useful in daily clinical practice where assessment of trainees is usually performed by only one rater. The D-study also confirmed that eight items was sufficient. Having included more items would have made the tool more complicated to use, and would add only very little reliability (Fig. 1). Actually, the use of several raters would reduce the number of needed items, but this could raise questions concerning content validity, and use of all eight items is recommended.

A good construct validity of an assessment tool is crucial, because it defines the ability of the tool to discriminate the different stages of experience, and thereby assess the stage and effect of training. This particular tool was able to discriminate between the three groups in a highly statistically significant way, which indicates good construct validity. Comparing the groups two by two showed statistical significance as well. This tool discriminates between novices and intermediates, which improves its usefulness in clinical practice. Previous studies with video assessment did not generally distinguish between more subtle differences in trainee performance [10], but this tool seems well-qualified for this purpose.

Most assessment tools are validated for formative use, which makes them useful to formal or informal feed-back in daily clinical practice. A validation for pass-fail tests (summative use) increases requirements because of higher consequences for the trainees. In this study we validated the tool for formative use. But as it is able to distinguish novices from intermediates, further specific validation for summative assessment use is possible with this tool.

The study has some limitations. We decided to use assessment of blinded video recordings to obtain less bias. Video recordings have inaccuracies in regard to camera angles and may not capture all elements of dissection. This was widely debated in our group but with the high quality pictures advantages of the method outnumbered disadvantages. Furthermore, we are aware of, that video recording will not always be possible in daily surgical practice. Blinded assessment may seem ideal if the rater must spend time both as attending surgeon and as rater. However, the tool is validated under the most ideal circumstances and may therefore be used blinded as well as un-blinded. It can be used for formative assessment of trainee's performance where attending surgeons are not required and reduce time spend for quality control.

We conducted a small study with ten participants and found consistent results. This tool is specifically designed for assessment of technical skills in Lichtenstein hernia repair and it may serve as a supplement to existing assessment tools. The advantages of our tool are: It is procedure-specific, have few (only eight) items, can be used for blinded rating, has high reliability which makes it useful with one rater, and has the ability to discriminate between surgical trainees. Technical skill is only one of many competencies required in today's surgery. Yet, it is a crucial skill. Lichtenstein hernia repair is a common procedure and one of the first surgical procedures encountered by young trainees; it is even mentioned as "sentinel procedure" [14]. This highlights the need for a valid and reliable assessment tool in this particular procedure.

Conclusions

Competence in Lichtenstein hernia repair could be assessed blindly in a reliable and valid fashion using this tool on video-recorded procedures. It could improve training outcome and ensure quality of surgical performance in surgical trainees. We recommend this procedure-specific tool for future assessment of trainees performing Lichtenstein hernia repair to ensure that the objectives of competency-based training in surgical skills are met.

Acknowledgments This research is founded by The Central Denmark Region, Aarhus University, Aalborg University Hospital and TRYG foundation. Thanks to participants involved in the development of the assessment form: Lars Maagaard Andersen, Finn Kallehave, Morten Gaarden, Hans Friis Andersen, and to doctors and patients accepting the video recording of the hernia repairs.

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Reznick RK, MacRae H (2006) Teaching surgical skills: changes in the wind. N Engl J Med 355:2664–2669
2. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C (2002) Shifting paradigms: from Flexner to competencies. Acad Med 77:361–367
3. Bell RH Jr, Biester TW, Tabuenca A, Rhodes RS, Cofer JB, Britt LD, Lewis FR Jr (2009) Operative experience of residents in US general surgery programs: a gap between expectation and experience. Ann Surg 249:719–724
4. Hald N, Sarker SK, Ziprin P, Villard PF, Bello F (2011) Open surgery simulation of inguinal hernia repair. Stud Health Technol Inform 163:202–208
5. Shulman A (1996) Lichtenstein Hernia repair and how to do them.. rigth! Wagner Design, California, ISBN 0-9653526-0-9
6. Amid PK, Shulman AG, Lichtenstein IL (1995) The Lichtenstein open "tension-free" mesh repair of inguinal hernias. Surg Today 25:619–625
7. Shulman AG, Amid P (1994) Which Lichtenstein method? Arch Surg 129:561
8. Shulman AG, Amid PK, Lichtenstein IL (1995) Mesh between the oblique muscles is simple and effective in open hernioplasty. Am Surg 61:326–327
9. Ghaderi I, Vaillancourt M, Sroka G, Kaneva PA, Vassiliou MC, Choy I, Okrainec A, Seagull FJ, Sutton E, George I, Park A, Brintzenhoff R, Stefanidis D, Fried GM, Feldman LS (2011) Evaluation of surgical performance during laparoscopic incisional hernia repair: a multicenter study. Surg Endosc 25:2555–2563
10. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J (2010) Objective assessment of technical surgical skills. Br J Surg 97:972–987
11. Faulkner H, Regehr G, Martin J, Reznick R (1996) Validation of an objective structured assessment of technical skill for surgical residents. Acad Med 71:1363–1365
12. Konge L, Lehnert P, Hansen HJ, Petersen RH, Ringsted C (2012) Reliable and valid assessment of performance in thoracoscopy. Surg Endosc 26:1624–1628
13. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B (2008) Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. BJOG 115:908–916
14. Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL (2005) Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. Surgery 138:640–647
15. Doyle JD, Webber EM, Sidhu RS (2007) A universal global rating scale for the evaluation of technical skills in the operating room. Am J Surg 193:551–555
16. Paisley AM, Baldwin P, Paterson-Brown S (2001) Feasibility, reliability and validity of a new assessment form for use with basic surgical trainees. Am J Surg 182:24–29
17. Downing SM, Yudkowsky R (2009) Assessment in health professions education. 1st edn. Routledge
18. Vogt VY, Givens VM, Keathley CA, Lipscomb GH, Summitt RL Jr (2003) Is a resident's score on a videotaped objective structured assessment of technical skills affected by revealing the resident's identity? Am J Obstet Gynecol 189:688–691
19. Al-Chalabi TS, Al-Na'ama MR, Al-Thamery DM, Alkafajei AM, Mustafa GY, Joseph G, Sugathan TN (1983) Critical performance analysis of rotating resident doctors in Iraq. Med Educ 17:378–384
20. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 84:273–278
21. Beard JD, Jolly BC, Newble DI, Thomas WE, Donnelly J, Southgate LJ (2005) Assessing the technical skills of surgical trainees. Br J Surg 92:778–782
22. Driscoll PJ, Paisley AM, Paterson-Brown S (2008) Video assessment of basic surgical trainees' operative skills. Am J Surg 196:265–272
23. Vassiliou MC, Feldman LS, Fraser SA, Charlebois P, Chaudhury P, Stanbridge DD, Fried GM (2007) Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. Surg Innov 14:211–216
24. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A (2008) Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. Ann Surg 247:372–379
25. Streiner DL, Norman GR (2008) Health measurement scales. 4th edn. Oxford, ISBN 978-0-19-923188-1