

ORIGINAL ARTICLE

Open Access



# Comparing self-assessment of laparoscopic technical skills with expert opinion for gynecological surgeons in an operative setting

Rami Kilani

## Abstract

**Background:** Competence in laparoscopic skills is important for all gynaecological surgeons. Most residency programmes teach technical skills in the operating room and through lectures, where the evaluation of surgical skills is usually done through subjective evaluation. After graduating residency, most surgeons depend on themselves to decide if they are competent in performing a certain procedure. The objective of this study is to evaluate the accuracy of surgeon self-assessment compared with expert assessment of competence in laparoscopic surgical skills. A double-blind prospective cohort study was undertaken at Prince Hamza Hospital between January 2016 and April 2016 in Amman, Jordan. Eight practicing gynecologists and obstetricians performed and recorded 88 laparoscopic procedures including ovarian cystectomy, salpingectomy for ectopic pregnancy, salpingoophorectomy, resection of endometriosis, adhesiolysis and ovarian drilling. Participating gynecologists recorded the procedures and were asked to complete a Global Rating Index of Technical Skills (GRITS) evaluation after the surgery testing across multiple areas with a lowest score of 8 and a highest score of 40. Two well-versed laparoscopic experts in objective structured assessment of technical skills (OSATS) also independently scored all procedures using the same parameters. The correlation coefficient and internal consistency were calculated.

**Results:** The GRITS score was calculated for each participant with a mean assessment score of 3.47 for each parameter. Participants self-assessment scores were significantly higher than expert assessment scores ( $p < 0.05$ ). The correlation coefficient was calculated and it can be seen that there was high inter-expert correlation in assessment across all participants evaluations ( $ICC > 0.90$ ).

**Conclusion:** Self-assessment of surgical laparoscopic skills is higher than expert evaluation of these technical skills. Quality assurance measures need to be revisited and restructured through more frequent assessments using peer and expert assessment alongside self-assessment. Gynecologists also need to undergo proper assessment prior to starting independently performing procedures that require new skills.

**Keywords:** Global Rating Index of Technical Skills (GRITS), Objective structured assessment of technical skills (OSATS), Intra-class correlation coefficient (ICC), Continuing Professional Development (CPD), Self-assessment, Surgical skills, Expert assessment

Correspondence: [rkilani2000@yahoo.com](mailto:rkilani2000@yahoo.com)  
Hashemite University, Zarqa, Jordan

## Background

Accurate self-assessment of knowledge and technical skills is essential for the safe and effective practice of medicine. Davis Boud defines self-assessment in his book “Enhancing Learning Through Self-Assessment” as the act of judging ourselves and making decisions about the next step. Boud’s opinion is that assessment can only be conducted against specific benchmarks or criteria [1]. Gordon published a systematic review on trainees from different health professions, college students, and graduate trainees [2]. The researchers explored self-assessment relative to an objective standard or an expert’s evaluation and concluded that self-assessment is fundamental to continuing medical competency and that self-assessment coupled with a specific set of criteria may lead to an improved outcome and more skilled professionals. Sullivan and Hall suggest that self-assessment promotes reflection on self-performance and motivates learners to react accordingly [3]. The application of the aforementioned concepts in medicine and specifically in laparoscopic surgery could lead to accurate self-assessment of performance by surgeons. This could eventually lead to the proper identification of strengths and weaknesses and allow the individual to create a plan for improvement.

Barnsley et al. looked at junior doctor self-assessment regarding confidence and competence of clinical skills versus objective assessment [4]. The researchers found no correlation between self-assessment and objective assessment. MacDonald et al. compared self-assessment of technical skills with simulator data in second and third year medical students with no previous exposure to laparoscopic training [5]. This selected task required the operator to pick up the target with one grasper and place it in the target box without releasing the target. Medical students were asked to evaluate their performance, and their evaluation was compared to the simulator data. The study found that self-assessment improves with repetition. Other researchers looked at resident self-assessment versus faculty assessment in performing laparoscopic procedures and found that residents were more critical of their performance than faculty members [6].

Arora et al. recently looked at self-assessment in technical and non-technical skills among surgical residents in a simulated environment [7]. Surgeons were asked to perform a laparoscopic cholecystectomy in a simulated laboratory. Two experts assessed the technical skills of surgeons, whereby the first expert watched the procedure live from a control room while the second expert evaluated technical skills after watching a video recording of the procedure. Both participants and experts used a validated objective tool. This study concluded that residents are accurate in self-assessment of their technical skills. However, that particular study and most literature examining this area of research face limitations to their

experiments including small sample sizes and the use of simulated procedures rather than actual procedures. According to the data that was collected, no evidence was found of a published study investigating self-assessment of laparoscopic technical skills of practicing gynecological surgeons performing specific procedures and comparing the evaluation to an external evaluator assessment. Thus, the aim of this study is to examine the use of self-assessment comparatively between gynecologists and experts with a large pool of participants performing laparoscopic procedures in the operative setting.

## Methods

### Examination process

Jordanian surgeons have adopted minimally invasive surgical techniques similarly to their counterparts in different areas of the world. Most Jordanian surgeons acquire the new skills through attending courses, workshops, and shadowing colleagues who have more experience in minimally invasive procedures. Privileges to perform surgeries are granted by the hospital based on qualifications. There is no official Jordanian recertification program after passing the specialty board exams, and the continuous medical education program is still at the early stages of development which makes a surgeon’s self-assessment of technical skills significantly more important.

The project is a prospective study. Participants are Jordanian board-certified obstetricians and gynecologists with privileges to practice at Prince Hamzah Hospital.

Candidates were approached by the primary investigator in the time period between January 2016 and April 2016 to participate in the study. They were supplied with an information leaflet explaining the research project, objectives, methods, and tasks involved. Surgeons who agreed to participate signed a consent form and were given a tutorial by the primary investigator on the Global Rating Index of Technical Skills (GRIT). This involved a 20-min session to familiarize them with the evaluation criteria and instructions on how to complete the evaluation form. An instruction sheet with all tutorial information was supplied to all participating surgeons. An example of a videotaped performance with predetermined scores was also shown to all participants.

Operative laparoscopic procedures including laparoscopic oophorectomy, laparoscopic ovarian cystectomy, laparoscopic salpingectomy, and adhesiolysis were evaluated. These procedures were chosen because they are the most common laparoscopic procedures performed at Prince Hamzah Hospital. Both gynecologists and external assessors were familiar with those procedures. The aforementioned procedures were also considered “short” procedures, thereby making the video assessment stage less time consuming.

The surgical lists at Prince Hamzah Hospital were reviewed the day prior to the surgery, and all participating gynecologists with operative laparoscopy cases were reminded to record the case and complete the GRIT. Emergency cases were also included, and the on-call gynecologist was asked to record the case and complete the form as well. The GRIT evaluation forms (Additional file 1) were available in all the gynecological operative suites. Extra copies were also available in a nearby office. The patients were not asked for permission to record the cases in accordance with the United Kingdom General Medical Council guidelines which states that a separate permission for recording a surgical procedure is not needed as long as the patient is anonymized.

Every participating gynecologist was assigned a number. Participating gynecologists recorded every procedure and were asked to complete the evaluation form after the surgery and included their assigned number and the procedure performed on every form. The form and DVD of the procedure were placed in personalized envelopes with the participant's number and collected daily.

The video recordings were sent to two of the external assessors who used the same GRIT to evaluate for technical skills. External assessors were blinded to which surgeon performed which procedure. The external assessors were experienced laparoscopic surgeons with experience in teaching and evaluating residents. They were familiar with the objective structured assessment of technical skill (OSAT) global rate scale. They scored separately and did not communicate during the scoring process.

### Statistical analysis

Data used for the descriptive statistics were obtained from intraoperative video records. Shapiro-Wilk test was used as test of normality. Mann-Whitney *U* test and Student *t* test were also used to test for distribution. Cronbach's alpha was used to calculate internal consistency. Internal consistency is a measure of reliability and measures whether several items that propose to measure the same general construct produce similar scores. Intra-class correlation coefficient (ICC) and Pearson correlation were used to measure inter-expert assessment reliability. Guidelines for evaluating the level of agreement among scores were >0.80 for excellent correlation, 0.60–0.80 for good correlation, 0.40–0.60 for fair correlation, and <0.40 for poor correlation.

### Results

A total of eight gynecologists met the inclusion criteria and agreed to participate in the study. Two surgeons were excluded due to the fact that they did not perform operative laparoscopy. The total number of procedures recorded during the study period was 88 cases. Ten cases were excluded; four recordings were incomplete and six were corrupted. This brought the total number

of procedures to 78 cases, collected from eight gynecologists. Qualifications and years of experience of the participating gynecologists can be seen in Table 1 which also highlights the number of surgical cases performed by the participating gynecologists. The videos varied widely in length from 8 min for an ovarian cyst aspiration procedure to 83 min for ablation of endometriosis procedure. The total length of all videos was 2655 min, and the mean was 37.4 min per video. The videos included mainly ovarian cystectomies, salpingectomies for ectopic pregnancy, salpingoophorectomies, resection of endometriosis, adhesiolysis, and ovarian drilling.

Internal consistency reliability was calculated for the GRIT without the communication skills, which were excluded due to the difficulty in observing this specific skill through video. Internal consistency reliability showed excellent reliability (Cronbach's alpha 0.883–0.904).

An initial descriptive analysis and normality test was carried out which indicated normal distribution of participants' scores with a Shapiro-Wilk normality test value of  $p > 0.05$ . In case of normal distribution, comparing means using *t* test is considered appropriate. The independent sample test was used since the populations were considered independent.

In Table 2, it can be seen that individual scores were evaluated significantly higher than expert evaluation ( $p < 0.05$ ). Figure 1 demonstrates the mean scores for each measured component for all participants, expert 1 and expert 2.

ANOVA test was carried out to determine whether there are any significant differences between the means of the total scores of all participants, expert 1 and expert 2 groups in Table 3. It was found that there was a statistically significant difference between the total scores of the participant and the experts; self-assessment was significantly higher than expert assessment.

Inter-expert assessment reliability was evaluated using ICC. All analyses of the inter-expert assessment reliability indicated excellent correlation as can be seen in Table 4. The total inter-expert assessment reliability for the two expert scores was calculated looking at the ICC and excellent reliability was noted (ICC = 0.9630).

### Discussion

Introspection and self-assessment are valuable traits for surgeons leading to comprehensive development of technical and personal skills. Overconfidence and lack of awareness of one's own abilities may lead to the inability to recognize limits and may endanger patients [8]. Self-assessment is thereby a significant measure of quality assurance that can potentially help improve patient safety and reduce error in the operating room.

Simulation learning has overtaken traditional methods in the training of new surgeons making self-assessment more important than ever. This type of teaching shifts learning

**Table 1** Participants' qualifications and years of experience

	Years of experience	Cases performed	Cases excluded	Cases used
Participant 1	6	14	2	12
Participant 2	8	10	-	10
Participant 3	7	12	1	11
Participant 4	9	8	2	6
Participant 5	18	7	1	6
Participant 6	20	13	2	11
Participant 7	12	13	2	11
Participant 8	10	11	-	11

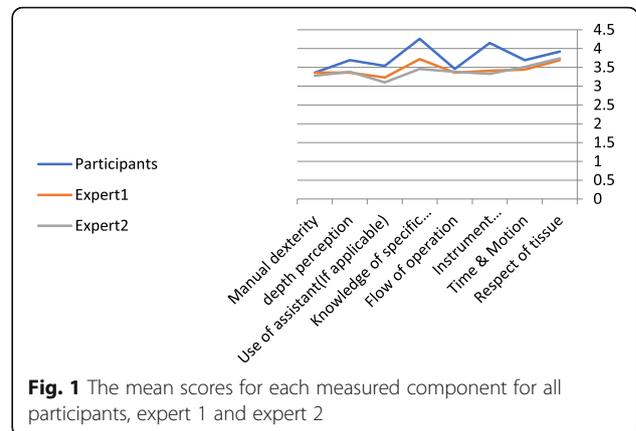
towards self-direction. Thus, surgeons must be able to accurately assess their abilities to personalize their training according to their individual performance [9]. Furthermore, self-assessment is an important parameter of personal development through continuous learning and has been shown to be an important part of a consultant's yearly appraisal [10].

New procedures requiring different technical skills are being introduced regularly in the field of minimally invasive surgery. Surgeons are thereby depending regularly on self-evaluation to determine if they can perform these procedures. Insufficient learning and inadequate evaluation of a surgeon's capabilities may lead to harming the patients. The advent of laparoscopic cholecystectomy led to surgeons offering the procedure to their patients after only attending one course leading to a major spike in common bile duct injuries. Researchers thereby explored the amount of experience needed to adequately perform a laparoscopic cholecystectomy and found it to be approximately 50 cases, owing to the fact that most complications occur in the first 30 cases [11].

Improved teaching and application of the new technology leads to decreased complications in laparoscopic procedures; however, surgeons are still solely responsible for

**Table 2** *t* test to compare the mean of participants and expert scores

The tested samples	<i>t</i> test for equality of means
All participants mean scores with expert 1 mean scores	<i>p</i> value 0.030
All participants mean scores with expert 2 mean scores	<i>p</i> value 0.017
All participants mean scores with the mean scores of both experts	<i>p</i> value 0.003



**Fig. 1** The mean scores for each measured component for all participants, expert 1 and expert 2

determining their competence in performing new procedures [12]. Our study demonstrates a lack of agreement between self-assessment and expert assessment of surgical technical skills thereby indicating that current self-assessment measures are inadequate. We used the Global Rating Index of Technical Skills (GRIT) (Additional file 1) for the self-assessment as well as the external evaluator assessment since this tool is documented and proven to be feasible, reliable, and valid [13].

Similar research to our study was done by Evans et al. and found that surgeons are not capable of effectively evaluating their technical surgical skills [8]. The authors compared self-assessment with peer assessment and expert assessment and concluded that surgeons tend to overestimate their technical skills. Comparatively, the participants in our study were more likely to overestimate their technical skills. In contrast, two recent studies from the Imperial College London [7, 10] found moderate to high correlation between self-assessment and an expert assessment for technical skills. However, the imperial study looked at students rather than independent practicing physicians.

The results of this study also suggest similar findings to Pandey et al. that demonstrated that surgeons may inaccurately self-assess their own skills and have difficulty accepting that their performance may be suboptimal [14]. The results of this study should encourage surgeons to enroll in formal assessment prior to starting to perform surgeries that require new skills. The results should also encourage surgeons to use self-assessment to improve their skills and to identify their strengths and work on

**Table 3** ANOVA test for the total scores of the participants and experts

ANOVA score					
	Sum of squares	df	Mean square	<i>F</i>	Sig.
Between groups	188.838	2	94.419	4.405	.014
Within groups	2443.282	114	21.432		
Total	2632.120	116			

**Table 4** Inter expert assessment reliability

Inter-expert assessment correlation	Intra-class correlation coefficient (ICC)
Participant 1 cases	0.962
Participant 2 cases	0.938
Participant 3 cases	0.975
Participant 4 cases	0.995
Participant 5 cases	0.975
Participant 6 cases	0.962
Participant 7 cases	0.938
Participant 8 cases	0.990

improving their weaknesses. Simulation labs would be an ideal environment to improve skills independently. This has been implemented in many different centers with promising results as shown in Arora et al. [7].

There is currently no true medical license recertification program in Jordan and several other countries in the Middle East and the rest of the world. Health officials and governing bodies need to consider developing a program or model for periodic evaluation of surgical skills encompassing cognitive and technical skills. To our knowledge, no systematic empirical research exist that measures self-assessment in an operative setting with qualified and experienced gynecologists with at least 7 years of experience. Our study is unique in the fact that it allows for measuring real-life technical skills of surgeons rather than assessment of simulation lab skills.

It is assumed that expert assessment is the gold standard as the best measure of evaluation. Many authors looked at verifying the claim that expert assessment is the best form of assessment when studying global rating scales [15]. They found contradicting results with most studies showing only moderate correlation between expert evaluation and raw scores [16]. It is difficult to find an alternative to expert assessment in medical education, which is why most studies, including this one, use experts as the measure for assessment.

The reliability of the expert assessment was also studied, and authors agree that experts are likely to agree among themselves given the chance to evaluate a short, structured, and simple task [17, 18]. Martin et al. also found high inter-reliability between experts watching a video recording of residents performing a standardized patient interview [9]. In this study, two experts scored the participants separately to improve the reliability of the expert assessment.

The limitations of our study include the use of means for comparison and sample size. Sample size in our study was limited since ethical approval was only obtained from Prince Hamzah Hospital in Amman; other hospitals and specialists thereby could not be included in the study. The comparison of group means as generated by the

participants and the experts may conceal individual differences [15]. This can also be seen in the research by Arnold et al. that examined subgroups of self-assessors. The authors found that high achievers tend to underestimate their performance while underachievers tend to overestimate their performance [19]. These findings may serve to reinforce the claim that there is a weak correlation between self-assessment scores and expert scores.

This study also focuses on the assessment of technical skills only. Non-technical skills are not evaluated since video recording of procedures is not considered a good tool for this type of evaluation. However, non-technical skills such as teamwork, leadership, situation awareness, decision making, task management, and communication are equally important, if not more important, than technical skills [20, 21]. Comparatively, studies that looked at self-assessment of non-technical skills showed that surgeons inadequately assess non-technical skills. It was also found that self-assessment of non-technical skills is significantly more overestimated when compared to expert evaluation with a sample of more experienced surgeons [7].

## Conclusions

This study shows that there remains a significant difference between self-assessment and expert assessment in the evaluation of laparoscopic technical skills for gynecological surgeons. Accurate self-assessment of technical skills in laparoscopy is important for practicing gynecologist as well as trainees to identify their strengths and weaknesses and improve their performance. Adequate self-assessment measures should encourage gynecologist to improve their skills independently in a simulated environment. This study showed that experienced gynecologist overestimated their surgical skills when compared to expert assessment. Quality assurance measures need to be revisited and restructured through more frequent assessments using peer and expert assessment alongside self-assessment. Gynecologists also need to undergo proper assessment prior to starting to independently perform procedures that require new skills.

## Additional files

**Additional file 1:** Represent the Global Rating Index of Technical Skills. (DOCX 317 kb)

**Additional file 2:** 2. Consent form. (DOCX 35 kb)

**Additional file 3:** 3. Participants information sheet. (DOCX 29 kb)

## Abbreviations

ABOG: American Board of Obstetrics and Gynecology; CPD: Continuing professional development; GOALS: Global Operative Assessment of Laparoscopic Skills; IUD: Intra uterine device; JMC: Jordan Medical Council; MOC: Maintenance of Certificate; MOH: Ministry of Health; OSATS: Objective structured assessment of technical skills; OSCE: Objective Structured clinical Exam; RCOG: Royal College of Obstetricians and Gynaecologists; RMS: Royal Medical Services

**Acknowledgements**

We owe our deepest gratitude to Dr. Mazen Fraij and Dr. Osama Badran, Consultant Gynecologists, who offered to score the videos and managed to fit this difficult task in their very busy schedules with no financial gain. We also want to thank Prince Hamzah Hospital staff for their outstanding contribution.

**Funding**

No external source of funding. First author sponsored all printed and visual materials.

**Availability of data and materials**

Study data and material are kept by the first author.

**Author's contributions**

RK designed the study, recruited the patients, analyzed the data, and wrote the paper. The author read and approved the final manuscript.

**Author's information**

Dr Rami Kilani is an Assistant Professor of Obstetrics and Gynecology at the Hashemite University-Zarqa, Jordan.

**Ethics approval and consent to participate**

Ethical approval was obtained from Prince Hamzah Hospital institutional review board. Surgeons who agreed to participate in the study signed a consent form (Additional files 2 and 3).

**Consent for publication**

Not applicable.

**Competing interests**

The author declares no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 5 January 2018 Accepted: 30 August 2018

Published online: 26 September 2018

**References**

- Boud D, Falchikov N (1995) What does research tell us about self assessment. In: Enhancing learning through self assessment. Kogan Page, London
- Gordon MJ (1991) A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 66(12):762–769
- Sullivan K, Hall C (1997) Introducing students to self-assessment. *Assess Eval High Educ* 22(3):289–305
- Barnsley L, Lyon PM, Ralston SJ, Hibbert EJ, Cunningham I, Gordon FC, Field MJ (2004) Clinical skills in junior medical officers: a comparison of self-reported confidence and observed competence. *Med Educ* 38(4):358–367
- MacDonald J, Williams RG, Rogers DA (2003) Self-assessment in simulation-based surgical skills training. *Am J Surg* 185(4):319–322
- Peyre SE, MacDonald H, Al-Marayati L, Templeman C, Muderspach LI (2010) Resident self-assessment versus faculty assessment of laparoscopic technical skills using a global rating scale. *Int J Med Educ* 1:37
- Arora S, Miskovic D, Hull L, Moorthy K, Aggarwal R, Johannsson H, Gautama S, Kneebone R, Sevdalis N (2011) Self vs expert assessment of technical and non-technical skills in high fidelity simulation. *Am J Surg* 202(4):500–506
- Evans AW, Leeson R, Petrie A (2007) Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery. *Med Educ* 41(9):866–872
- 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(2):273–278
- Moorthy K, Munz Y, Orchard TR, Gould S, Rockall T, Darzi A (2004) An innovative method for the assessment of skills in lower gastrointestinal endoscopy. *Surg Endosc Interv Tech* 18(11):1608–1619
- Ellison EC, Carey LC (2008) Lessons learned from the evolution of the laparoscopic revolution. *Surg Clin N Am* 88(5):927–941
- Luchtefeld M, Kerwel TG (2012) Continuing medical education, maintenance of certification, and physician reentry. *Clin Colon Rectal Surg* 25(3):171
- 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(5):551–555
- Pandey VA, Wolfe JHN, Black SA, Cairois M, Liapis CD, Bergqvist D (2008) Self-assessment of technical skill in surgery: the need for expert feedback. *Ann R Coll Surg Engl* 90(4):286–290
- Ward M, Gruppen L, Regehr G (2002) Measuring self-assessment: current state of the art. *Adv Health Sci Educ* 7(1):63–80
- Risucci DA, Tortolani AJ, Ward RJ (1989) Ratings of surgical residents by self, supervisors and peers. *Surg Gynecol Obstet* 169(6):519–526
- Regehr G, MacRae H, Reznick RK, Szalay D (1998) Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 73(9):993–997
- Miller JD, McCain J, Lynam DR, Few LR, Gentile B, MacKillop J, Campbell WK (2014) A comparison of the criterion validity of popular measures of narcissism and narcissistic personality disorder via the use of expert ratings. *Psychol Assess* 26(3):958
- Arnold L, Willoughby TL, Calkins EV (1985) Self-evaluation in undergraduate medical education: a longitudinal perspective. *Acad Med* 60(1):21–28
- Salas E, Bowers CA, Edens E (eds) (2001) Improving teamwork in organizations: applications of resource management training. CRC press/Taylor and Francis group. ISBN 9780805828450 - CAT# ER5855
- Siu J, Maran N, Paterson-Brown S (2016) Observation of behavioural markers of non-technical skills in the operating room and their relationship to intra-operative incidents. *Surgeon* 14(3):119–128

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)