

Objective assessment of obstetrics residents' surgical skills in caesarean: Development and evaluation of a specific rating scale

Quentin Berl, Noémie Resseguier, Maria Katsogiannou, Franck Mauviel, Xavier Carcopino, Léon Boubli, Julie Blanc

▶ **To cite this version:**

Quentin Berl, Noémie Resseguier, Maria Katsogiannou, Franck Mauviel, Xavier Carcopino, et al.. Objective assessment of obstetrics residents' surgical skills in caesarean: Development and evaluation of a specific rating scale. Journal of Gynecology Obstetrics and Human Reproduction, 2021, 50 (1), pp.101812. 10.1016/j.jogoh.2020.101812 . hal-03149992

HAL Id: hal-03149992
https://hal.science/hal-03149992

Submitted on 9 May 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.

Objective assessment of obstetrics residents' surgical skills in caesarean: Development and evaluation of a specific rating scale

Quentin Berl[a], Noémie Resseguier[b], Maria Katsogiannou[c], Franck Mauviel[d], Xavier Carcopino[a,e], Léon Boubli[a], Julie Blanc[a,b,*]

[a] Department of Obstetrics and Gynecology, Nord Hospital, APHM, Chemin des Bourrely, 13015, Marseille, France
[b] EA 3279, Public Health, Chronic Diseases and Quality of Life, Research Unit, Aix-Marseille University, 13284, Marseille, France
[c] Hôpital Saint Joseph, Department of Obstetrics and Gynecology, FR-13008, Marseille, France
[d] Department of Obstetrics and Gynecology, Ste Musse Hospital, 54, rue Henri Sainte Claire Deville, 83000, Toulon, France
[e] Aix-Marseille University (AMU), Univ Avignon, CNRS, IRD, IMBE UMR, Marseille, France

A B S T R A C T

Objective: To develop a modified version of Objective Structured Assessment of Technical Skill (OSATS) rating scale for evaluation of surgical skills specific to caesarean and to assess its relevance in documenting the residents' learning curve during their training. Secondarily, to verify the scale's stability to caesarean's level of difficulty and comparing self-assessment to hetero-assessment in order to propose a practical application of this rating scale during residency.
Study Design: We conducted a multicentre observational prospective study, from May 2018 to November 2018. All residents at that time could participate and fill in the rating scale after caesarean. Senior surgeons had to fill in the same rating scale. We analysed correlation between self-assessments and hetero-assessments and sensitivity to change of the rating scale. Analysis of feature's relevance was performed by principal component analysis, factor analysis and reliability analysis.
Results: In total, 234 rating scales were completed evaluating 18 residents. Our study demonstrated that our rating scale could be used to evaluate surgical skills of residents during caesarean and distinguish their year of residency ($p < 0.001$) with a high correlation between self and hetero-assessment (Intraclass Correlation coefficient for global score: 0.78; 95% CI 0.68−0.86). The principal component analysis revealed two dimensions corresponding to the two parts of the rating scale and the factorial analysis allowed us to confirm distribution of features according to these two dimensions. Cronbach's alpha allowed us to highlight the percentage of representation of the scale's features in relation to all potential theoretical features (0.93, 95% CI 0.82−0.95).
Conclusion: Our rating scale could be used for self-assessment during residency and as a hetero-assessment tool for validating defined stages of the internship.

Introduction

Learning the technique of surgery is central in every surgical residency program. To gain autonomy, the resident has to have theoretical knowledge, technical dexterity, communication skills, leadership and to be able to make decisions [1]. For a few years, new teaching and evaluation approaches have been developed [2]. Assessment methods have to be objective, standardized, feasible, validated, reliable, practical and inexpensive [3]. The senior surgeon have to maintain control by gradually reducing supervision as the resident gains surgical skills [4]. In 2005, the Royal College of Obstetricians and Gynaecologists (RCOG) defined six surgical procedures that need specific assessment procedures: caesarean section (CS), instrumental delivery, perineal repair, uterus revision, hysterectomy and exploratory laparoscopy [5].

CS is the most frequently performed conventional open surgical procedure in the world with 29.7 millions of births per year; it represented 20.4% of deliveries in France in 2016 [6]. This surgery is performed in a repeated and precocious manner by obstetrics and gynaecology (OB/GYN) residents [7]. Several surgical skills can be acquired by simulation software [8] or simulators [9,10], educational videos [11], or even anatomic dissection [10,12] but all these methods are insufficient for residents training [13]. In 2007, the RCOG officially used an educational tool called "objective

structured assessment of technical skills" (OSATS) [14,15]. This model, initially created for simulation, was modified and adapted to CS in 2005 by the RCOG [5]. It is composed of a first part when the assessor determines whether the resident needs help in a series of key steps of CS, called "task list". The second part evaluates the resident's overall surgical skills ("global rating scale") [16,17]. This surgical learning assessment has shown excellent external validity in several studies [15,17–19]. The validity of OSATS scale adapted to caesarean has been studied only by reporting trainees' experiences [20].

In France, the resident's logbook is rarely used [7]. There is an actual request from residents for a formative assessment of theoretical knowledge and weighted practical assessments [7]. Unlike RCOG, there is no assessment scale of CS difficulty in France. Residents self-assessment is a current concern [21–24].

Our primary objective was to develop an original surgical skills assessment scale adapted to CS and integrating a task list with a Likert scale to use by self and hetero-assessment and to evaluate its relevance in documenting the progression of resident's scores with the year of residency. Secondary objectives consisted in checking the scale's stability to caesarean section difficulty and comparing self-assessment to hetero-assessment in order to propose a practical application of this rating scale during internship.

Material and methods

Study design and data collection

This prospective observational multicentre study was conducted from May 2018 to November 2018 in three university hospitals where residents could perform CS with a senior surgeon from the first year of residency. All residents present at that time could participate and fill-in the rating scale right after CS whatever gestational ages (GA). Senior surgeons had to fill in the same rating scale. The ones that were not completed by both residents and senior doctors were excluded from our analysis.

The Aix-Marseille University Ethics Committee approved this study (N°2017-14-12-003).

The criteria related to CS difficulty and its characteristics were collected: time of procedure, GA, number of uterine scars, cervix dilation, body mass index (BMI), characteristics of CS (type of emergency code), total estimated blood loss, multiple pregnancy, clotting disorders, placentation abnormalities, general anaesthesia. Data were collected and anonymised. A composite end-point of preoperative characteristics was defined by: BMI $\geq$ 30 or full dilation or gestational age < 28 weeks or uterine scar $\geq$ 1 or placenta praevia or accreta or polymyomatous uterus or clotting disorders.

Rating scale

The rating scale included the OSATS scale used by RCOG for CS. Several elements of global rating scale (GRS) adapted to obstetrics according to recent publications were added to the rating scale. [15,25,26] The relevance of the scale's content was verified and validated by three experts (JB, LB, XC) with more than 10 years of experience in OB/GYN. The scale was divided in two parts. The first part assessed resident autonomy for each surgery procedure step and the second part evaluated overall procedure of technical and non-technical skills [14,15,26,27].

Autonomy score – surgery procedure steps

The first part was worth maximum 70 points and reflected the resident's autonomy analysis during the different procedure steps with a five-point Likert scale [28,29]. Likert scale is more compelling than "pass/fail" [29]. "0" was given in case the senior surgeon decided to perform the CS by himself. All Likert scores for each procedure step were then added together in order to have a global evaluation of autonomy. Features were the following ones: appropriate skin incision (e.g. length, position), safe entry of peritoneal cavity, careful management of bladder, appropriate uterine incision (e.g. length, position), safe and systematic delivery of baby, appropriate delivery of placenta, check uterine cavity (e.g. intact, empty, configuration), safe securing of uterine angles, check for ovarian pathology, appropriate closure of rectus sheath, attention to haemostasis, neatness of skin closure, appropriate management of deeply engaged head.

Technical and behavioural score

The second part assessed the resident's overall technical skills such as tissue respect, movement fluidity, long-term planning, suture technique, and management of operational assistance. Subjective aspects were also collected such as involvement in management, relational attitude with the patient and the surgical team, adaptation to emergency and/or to situation. Total score was worth a maximum of 35 points with a five-point Likert scale for each assessed item. Features are detailed below: (i) respect for tissue, (ii) time, motion and flow of operation and forward planning, (iii) suturing and knotting skills as appropriate for the procedure, (v) relations with patient and the surgical team, (vi) insight/attitude, (vii) documentation of procedures.

Global score

Global score (105 points maximum) was the sum of first- and second-part assessment scores.

Global subjective assessment of the procedure's difficulty was marked from 1 to 5. A final feature allowed the senior surgeon to give a global evaluation of the resident's behaviour (1–4 points); "simple supervision; passive assistance; active assistance; shows and explains" (Fig. 1).

Assessment scale validity

We analysed the distribution of global self- and hetero-assessment scores according to year of residency (from 1st to 5th year) and to pre- and intraoperative characteristics: "English three-point score" (1=first caesarean or one scar with longitudinal incision, 2=twins and preterm before 28 weeks of gestation and two scars and obesity, 3=preterm before 28 weeks and placenta praevia and accreta), BMI $\geq$ 30, full dilation, CS urgency, postpartum haemorrhage (PPH) $\geq$ 500 mL, subjectivity level from 1 to 5, composite criterion of preoperative CS characteristics: BMI $\geq$ 30 or full dilation or GA < 28 weeks or uterine scar $\geq$1 or placenta praevia or placenta accreta or polymyomatous uterus or clotting disorders.

Statistical analysis

Statistical analysis was performed using RStudio V.1.0.44. We first carried out a descriptive analysis. Qualitative variables were expressed as absolute values and percentages. Quantitative variables were expressed as mean and standard deviation. In order to evaluate correlation between self-assessments and hetero-assessments, we calculated an intra-class correlation coefficient (ICC) with its 95% confidence interval (CI) for: autonomy score, technical and behaviour score, global score, subjective evaluation of CS difficulty and senior surgeon's behaviour. Sensitivity to change of the rating scale was analysed by comparing total scores according to: residents' year of residency, preoperative characteristics ("English score"), BMI $\geq$ 30, full dilation, occurrence of PPH, subjectivity level, composite criterion of caesarean section characteristics (BMI $\geq$ 30

or full dilation or GA < 28 weeks or uterine scar ≥1 or placenta praevia or placenta accreta or polymyomatous uterus or clotting disorders), emergency of CS defined by red code alert. Comparisons were done for binary variables with Student t test when appropriate, or with Mann-Whitney U test. For categorical variables, comparisons were done with the Kruskal–Wallis test. Analysis of feature's relevance was performed by principal component analysis for identification of the number of

A

EVALUATION CESARIENNE

Remplir **2 évaluations** par intervention : 1 par l'interne **ET** 1 par le Senior

	OUI	NON
Placenta Ant Recouvrant		
Placenta *Accreta*		
Jumeaux		
Ut. Poly-myomateux		
Troubles de la coagulation		
Anesthésie générale		

Contexte à remplir:

Nombre de cicatrices utérines :
Code de la césarienne : VERT / ORANGE / ROUGE Dilatation en CM :
Terme en SA :
IMC :

Temps opératoire (Incision / Fermeture) : Pertes estimées en ml :

Cocher la proposition la plus juste pour chaque étape :
Répondre "non" si le senior réalise l'étape en tant qu'opérateur :

	Est-ce que l'étape est réalisée par l'interne ?					
Etapes étudiées	**Non**	**Non, malgré indications**	**Partiellement avec aide**	**Oui, avec instructions**	**Oui, avec instructions minimes**	**En autonomie**
Connaissance de : la position fœtale, Hb, position du placenta						
Mise en place de la table opératoire						
Check liste et dire « incision »						
Incision cutanée						
Incision de l'aponévrose et accès intra-péritonéal						
Sécurisation de la vessie						
Hysterotomie						
Extraction fœtale						
Délivrance dirigée et révision utérine						
Hystérorraphie avec points d'angles						
Vérification de l'absence de pathologie annexielle						
Fermeture aponévrotique après compte des compresses						
Fermeture cutanée						
Consignes post opératoire						

Pour chaque domaine répondre de manière globale au sujet de l'interne Entourer la case la plus appropriée à votre perception

Fig. 1. Rating scale.

B

Implication dans la prise en charge						
Non connaissance du dossier Non compréhension de l'indication	**1**	**2**	**3**	**4**	**5**	Synthèse pertinente du dossier Anticipation de la prise de décision Evaluation claire des différentes stratégies
Respect du tissu						
Force inutile sur les tissus Utilisation inappropriés des instruments	**1**	**2**	**3**	**4**	**5**	Instrumentation maîtrisée Bonnne gestion des compresses
Temps, fluidité du mouvement, planification à long terme						
Mouvements inutiles Actions fréquemment arrêtées et rediscutée par le senior	**1**	**2**	**3**	**4**	**5**	Économie de mouvement Efficacité maximale Anticipation et enchainement des temps opératoires sans effort
Technique de suture						
Suture mal positionnée, nœud mal serré ou geste non sécurisé Erreure dans le choix des fils	**1**	**2**	**3**	**4**	**5**	Sutures positionnées de façon cohérente Nœuds sûrs Respect des règles de sécurité
Gestion des aides opératoires (interne = opérateur ==> senior + externe = aide op)						
Mal positionnées ou n'aidant pas Absence d'interaction	**1**	**2**	**3**	**4**	**5**	Potentialisation des assistants Informations claires et pertinentes
Relation avec la patiente et l'équipe chirurgicale						
Mauvaise communication Non perception des attentes du patient et/ou de l'équipe	**1**	**2**	**3**	**4**	**5**	Communication et agissements selon besoins du patient et/ou de l'équipe
Adaptation au degré d'urgence et/ou à la situation						
Déstabilisé Perte de contrôle	**1**	**2**	**3**	**4**	**5**	Bonne capacité d'adaptation Garde une vision d'ensemble

EVALUATION CESARIENNE 2eme PARTIE

Comportement du sénior lors de l'intervention :	**Montre et explique**	**Aide active**	**Aide passive**	**Supervision simple**

Degré de difficulté de la césarienne (anatomie et contexte)						
Simple	**1**	**2**	**3**	**4**	**5**	Très difficile

Fig. 1. (Continued)

dimensions in the rating scale. Factorial analysis then allowed to present distribution of features according to the identified dimensions and to assess factor loadings in order to retain only the most relevant ones. Finally, according to the feature universe model, we estimated the Cronbach's alpha coefficient, and the 95% CI was obtained by bootstrap method. In order to evaluate the relevance of the "subjective difficulty of CS", we described and compared values of this quantitative variable according to the composite criterion of CS characteristics, occurrence of PPH and red code alert with a Mann–Whitney U test. All tests were two-tailed. $P < 0.05$ was considered statistically significant.

Results

Study design and population

During the period of our study, 886 CS (of which 27.76% were elective) were performed in three maternity units where 25 residents were present. Among the 167 caesarean deliveries assessed, we excluded from our analysis 23 rating scales filled in only by residents and 12 filled in only by senior surgeons. Fifteen rating scales were also excluded because filled in by a resident who decide to change medical speciality from OB/GYN to psychiatry during the study time.

A total of 234 assessments by residents and senior surgeons (117 CS) were analysed. The majority of the assessments, 92 (39.31%) were carried out by a 5th-year resident; three 4th-year residents carried out 58 assessments (24.78%); four 3rd-year and six 1st-year residents carried out respectively 36 assessments (15.38%); finally, four 2nd-year residents carried out only 12 assessments (5.13%). Among the 18 evaluated residents, 1 (5.55%) was a 5th-year resident, 3 (16.67 %) by a 4th-year resident, 4 (22.22%) by a 3rd-year resident, 4 (22.22%) by a 2nd-year resident and 6 (33/33%) by a 1st-year resident.

Among the 234 assessments, 60 (25.64%) and 68 (29.06%) were performed in two tertiary care level maternities, and 108 (46.15%) in one secondary care level maternity.

Among the 24 senior surgeons, one (4.17%) was a university professor, 14 (58.33%) were hospital practitioners and 9 (37.50%) were clinic head assistants.

Characteristics of evaluated caesarean deliveries

CS characteristics are presented in Table 1.

Analysis of self-assessment and hetero-assessment rating scales

Mean scores for different parts of the modified OSATS scale as well as total scores were calculated (Table 2). Global mean score of the two parts (on 105 points) was 95.08 (± 12.30) for self-assessments and 97.48 (± 11.01) for hetero-assessments. ICC evaluated correlation between residents and senior surgeons scores for the different assessed factors. We observed ICC > 0.5 for CS difficulty, senior surgeon behaviour, first and second parts scores as well as global score.

Evaluation according to year of residency

We observed progression of total self- and hetero-assessment scores with year of residency (Fig. 1). Indeed, 1st-year residents obtained lower assessment scores than 4th and 5th-year residents (Table 3). Our analysis showed a statistically significant difference ($p < 0.001$) among mean scores, which varied increasingly with the year of residency.

Table 1
Characteristics of evaluated caesareans.

Characteristics	Results
Time of procedure[a] (min)	36.14 ± 16.63
Total estimated blood loss[a] (mL)	297.43 ± 265.70
• Post-Partum Haemorrhage[b]	10 (8.55%)
Gestational age[a] (weeks)	37.48 ± 4.34
• Gestational age < 28 weeks[b]	7 (5.98%)
Body Mass Index (BMI)[a] N=66	28.51 ± 6.15
• BMI ≥ 30[b]	30 (45.45%)
Number of uterine scars[a]	0.63 ± 0.91
Cervix dilatation[a] (cm)	2.62 ± 3.10
• Full dilatation[b]	7 (5.98 %)
Caesarean	
• Elective[b]	35 (29.91%)
• Green code[b]	43 (36.75%)
• Orange code[b]	34 (29.06%)
• Red code[b]	5 (4.27%)

Post-Partum Haemorrhage defined by total estimated blood loss ≥.500 mL.
Colour code defined by expected decision-birth duration: green > 30 min, orange <30 min, red < 15 min.
[a] Quantitative variables: Mean ± Standard Deviation.
[b] Qualitative variables: n (%).

Assessment scores progression according to CS difficulty

We did not observe any significant difference in the mean assessment score according to the "English" score (p = 0.51), full dilation (p = 0.07), preoperative composite criterion (see Methods section) (p = 0.71) and CS code (p = 0.29) (Table 4). Impact of BMI on assessment score was on the borderline of significance (p = 0.05). Mean scores were quite close for patients with BMI ≥ 30 (97.32 ± 9.88) and those with BMI < 30 (96.55 ± 12.83). Occurrence of PPH during CS was statistically associated with assessment scores (p < 0.001). Finally, subjectivity level (based on senior surgeon's and resident's perception) was statistically associated to assessment scores (p = 0.002): the higher was the subjectivity level, the lower the assessment score was.

Subjective difficulty of CS

Subjective difficulty score was significantly associated with the composite criterion of preoperative characteristics (mean subjective score 2.8 ± 1.04 in case of composite criterion vs 1.89 ± 0.98 when non composite criterion, p < 0.001) and PPH (mean subjective score 3.6 ± 1.14 in case of PPH vs 2.38 ± 1.04 in absence of PPH, p < 0.001).

Rating scale features' relevance

The principal component analysis revealed two dimensions, two logical sets of features, corresponding to the two parts of the rating scale. Then, factorial analysis of the different stages of CS allowed us to confirm distribution of features according to these two dimensions and to sort them according to their factor loadings. Skin suture was the feature with the lowest factor loading (0.29). Several aspects of the factorial analysis obtained by senior surgeons' and residents' assessments were similar.

Cronbach's alpha allowed us to highlight the percentage of representation of the scale's features in relation to all potential theoretical features (0.93, 95% CI 0.82−0.95).

Discussion

Our study showed that this rating scale is a reliable tool that could be used to evaluate surgical skills of residents during CS according to their year of residency and with a high correlation between self and hetero-assessment. This suggests the relevance of our scale in assessing residents and according to the CS level of difficulty and its different procedure steps.

The difference in scores was more significant among first-year residents and 5th-year residents in agreement with other previously published studies. Unlike other studies [14,29,29,30], we chose to integrate a task list by evaluating resident's autonomy during each procedure step of CS by using a Likert scale. We believe that in an educational objective, it is interesting to highlight the procedure steps that require special attention while pointing out the simple ones that can be carried out independently.

We revealed that the degree of emergency of CS was not statistically related to the total score unlike PPH. Indeed, in case of PPH, the senior surgeons usually give less autonomy to the residents since time is the key of the management of PPH. The composite criterion was not statistically related to the total score probably because lack of power. Some variable of the composite criterion are at limit of significance (BMI ≥ 30 and full dilatation). If demonstrated by a larger study, the features of the composite

Table 2
Correlation between self and hetero-assessment rating scales. Mean ± Standard Deviation.

Parts of rating scale	Self-assessment scores N = 117	Hetero-assessment scores N = 117	ICC (95% CI)
Subjective evaluation of caesarean difficulty	2.33 ± 1.08	2.63 ± 1.11	0.71 (0.61–0.80)
Senior surgeon's behaviour	3.02 ± 1.01	3.12 ± 0.86	0.69 (0.52–0.80)
First Part = Autonomy score	64.05 ± 8.91	64.84 ± 8.94	0.83 (0.71–0.91)
Second Part = Technical and behaviour score	31.03 ± 4.42	32.64 ± 2.79	0.55 (0.46–0.64)
Global score	95.08 ± 12.3	97.48 ± 11.01	0.78 (0.68–0.86)

ICC, Intraclass Correlation Coefficient; CI, Confidence Interval.

Table 3
Mean global scores depending on year of residency and caesarean characteristics.

Year of residency	Mean global scores	*p*
1	80.61 ± 17.44	*<0.001*
2	91.92 ± 5.15	
3	95.69 ± 8.94	
4	98.66 ± 7.67	
5	101.71 ± 5.5	
Caesarean characteristics	Mean global scores	*p*
« English » Score		
1	97.06 ± 9.50	0.51
2	96.62 ± 11.66	
3	90.45 ± 18.93	
BMI		
< 30	96.55 ± 12.83	0.053
≥ 30	97.32 ± 9.88	
Full dilatation		
No	96.39 ± 11.88	0.07
Yes	94.57 ± 8.87	
PPH		
No	97.70 ± 10.08	<0.01
Yes	81.10 ± 16.66	
Subjectivity level		
1	97.24 ± 9.66	<0.01
2	99.45 ± 7.39	
3	94.89 ± 14.28	
4	91.68 ± 15.00	
5	89.20 ± 13.75	
Preoperative composite criterion		
No	96.48 ± 11.92	0.71
Yes	96.17 ± 11.63	
Caesarean colour code		
Red	92.50 ± 12.12	0.29
Orange	96.01 ± 10.77	
Green	96.63 ± 12.10	

Mean ± Standard Deviation.
BMI: Body Mass Index.
PPH: Post-Partum Haemorrhage defined by estimated blood loss ≥ 500 mL.
Colour code defined by expected decision-birth duration: green > 30 min, orange <30 min, red < 15 min.

criterion (BMI ≥ 30 or full dilation or gestational age < 28 weeks or uterine scar ≥ 1 or placenta praevia or accreta or polymyomatous uterus or clotting disorders) could be considered to select CS for self-assessment or hetero-assessment.

Even though filling-in the rating scale took less than 3 min, it was not mandatory, and this could explain why all CS performed during the study period were not evaluated. The proportion of evaluated CS is low as previously reported in a study dealing with the assessment of instrumental deliveries [31]. This could be explained by the busy activity in the three maternity units most of the time. Furthermore, in accordance with Gosman et al. [32], we believe that the lack of surgical assessment was not related to the resident's performance, but rather to the constraint that may represent, for some, the assessment process.

First limitation of our study is the duration of our study (6 months), not long enough to allow progression follow-up of residents over time. Secondly, we reported an unequal repartition of the residents depending on their year of residency. This could be explained by the distribution of the residents in the three obstetrics units during only one semester. This could be avoided by increasing the length of the study. Thirdly, it was impossible to conduct a blinded study. This bias could be avoided by an external independent assessor or by videos of surgical procedure sequences. Nevertheless, Goff et al. demonstrated that surgical skills could be assessed with equally strong validity and reliability regardless of whether administration of OSATS scales was blinded or not [33]. Several authors suggest that OSATS are biased [34]. This can be avoided by increasing the sample size.

Our study presents several strengths, including its originality and its prospective and multicentre design. To our knowledge, no other study has been conducted for evaluation of OB/GYN residents on their overall surgical skills of CS, by self- and hetero-assessment. Similarities of the factorial analysis for the self and hetero-assessment are an additional proof of the conceptual validity of our tool. In addition, our study is the first to look at the progression of surgical evaluation scores according to CS clinical difficulty levels.

Direct observation of performance through specific assessment criteria was considered to be a valid and reliable method for evaluating the residents' surgical skills [14,15,35,36]. Self-assessment is generally approached in relation to the concept of autonomy as an aim of the educational process. Colthart et al. gave an operational definition to self-assessment as being a "personal evaluation of ones' characteristics and professional abilities in relation to perceived standards" [37]. Several authors demonstrated high validity and reliability of the OSATS scale [15,35,38,39]. The correlation between self and hetero-assessment was high for the global score but the ICC was between 0.5 and 0.75 for the technical and behaviour score. This could be explained by the usual underestimation in case of self-assessment. The parallel use of our assessment scale by residents and senior surgeons may also serve as a basis for discussion and could facilitate the feedback given by the senior surgeon. By revising the different features of the rating scale together, senior surgeons could reveal potential overvaluation errors, leading to improvement of the resident's self-assessment ability [40]. We think that this mode of assessment with Likert scales is a better way to respect educational process and accompany the residents than defining a threshold for each part to confirm the autonomy of the residents in performing a CS. However, further studies could evaluate the relevance of a threshold for the global score to confirm the autonomy of residents in performing a CS in pre-defined clinical situations at each stage of residency.

Our assessment scale could be used for self-assessment during the whole residency training. In addition, hetero-assessment could be validating at the end of first-year as well as the fourth and fifth-years of residency. Pre- and intraoperative criteria would allow to « select» the appropriate surgical situations to assess each level. These findings could be confirmed by evaluating this assessment scale is a larger sample of residents.

Funding

None

Declaration of competing interest

The authors declare that they have no conflict of interest.

Contributors to authorship

Q Berl: Project development, Data collection, Manuscript writing
N Rességuier: Data analysis, Data management
M Katsogiannou: Data management, Manuscript writing
F Mauviel: Data collection
X Carcopino: Project development
L Boubli: Project development
J Blanc: Project development, Data management, Data analysis, Manuscript editing

Details of ethics approval

This study was approved by the Aix-Marseille University Ethics Committee (N°2017−14-12−003).

Informed consent statement

Informed consent was obtained from all individual participants (residents and senior) included in the study.

Acknowledgements

We thank Michelle Meunier for helping to correct the English.

References

[1] Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. BMJ 2003;327(October 7422):1032–7.
[2] Rodriguez-Paz JM, Kennedy M, Salas E, Wu AW, Sexton JB, Hunt EA, et al. Beyond 'see one, do one, teach one': toward a different training paradigm. Postgrad Med J 2009;85(May 1003):244–9.
[3] Desender LM, Van Herzeele I, Aggarwal R, Vermassen FEG, Cheshire NJW. Training with simulation versus operative room attendance. J Cardiovasc Surg (Torino). 2011;52(February 1):17–37.
[4] Müller I, Zimmermann R. [The learning curve in the context of the cesarean section]. Gynakol Geburtshilfliche Rundsch 2003;43(October 4):238–44.
[5] Bisson DL, Hyde JP, Mears JE. Assessing practical skills in obstetrics and gynaecology: educational issues and practical implications. Obstet Gynaecol. 2006;8(April 2):107–12.
[6] Boerma T, Ronsmans C, Melesse DY, Barros AJD, Barros FC, Juan L, et al. Global epidemiology of use of and disparities in caesarean sections. Lancet Lond Engl. 2018;392(10155)1341–8 13.
[7] Rathat G, Hoa D, Gagnayre R, Hoffet M, Mares P. Formation chirurgicale des internes, spécialistes en gynécologie-obstétrique : résultats d'une enquête électronique nationale. J Gynécologie Obstétrique Biol Reprod. 2008;37 (November 7):672–84.
[8] Bouin T. La simulation sur mannequin haute fidélité dans la formation initiale des sages-femmes. Rev Sage-Femme. 2013;12(April 2):93–7.
[9] Tassin M, Cordier A-G, Laher G, Benachi A, Mandelbrot L. Simulateur d'amniocentèse : intérêts et développement d'un nouveau modèle reproductible et économique. J Gynécologie Obstétrique Biol Reprod. 2012;41 (November 7):679–83.
[10] Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, et al. Assessment of technical skills transfer from the bench training model to the human model. Am J Surg 1999;177(February 2):167–70.
[11] Driscoll PJ, Paisley AM, Paterson-Brown S. Video assessment of basic surgical trainees' operative skills. Am J Surg 2008;196(August 2):265–72.
[12] Sarker SK, Patel B. Simulation and surgical training. Int J Clin Pract 2007;61 (December 12):2120–5.
[13] Alici F, Buerkle B, Tempfer CB. Objective Structured Assessment of Technical Skills (OSATS) evaluation of hysteroscopy training: a prospective study. Eur J Obstet Gynecol Reprod Biol 2014;178(July):1–5.
[14] Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84(February 2):273–8.
[15] Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg 1997;173(March 3):226–30.
[16] Swanson DB, van der Vleuten CPM. Assessment of clinical skills with standardized patients: state of the art revisited. Teach Learn Med 2013;25 (Suppl 1):S17–25.
[17] Goff B, Mandel L, Lentz G, Vanblaricom A, Oelschlager A-MA, Lee D, et al. Assessment of resident surgical skills: is testing feasible? Am J Obstet Gynecol 2005;192(April 4)1331–8 discussion 1338-1340.
[18] Larsen C, Grantcharov T, Schouenborg L, Ottosen C, Soerensen J, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. BJOG Int J Obstet Gynaecol. 2008;115(June 7):908–16.
[19] van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. Br J Surg 2010;97(July 7):972–87.
[20] Objective Structured Assessment of Technical Skill in assessing technical competence to carry out caesarean section with increasing seniority. Best Pract Res Clin Obstet Gynaecol 2013;27(April 2):197–207.
[21] Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Acad Med J Assoc Am Med Coll. 1991;66(December 12):762–9.
[22] Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art. Adv Health Sci Educ Theory Pract 2002;7(1):63–80.
[23] Bresson L, Devos P, Maillet A, Verbecq P, Ardaens Y, Dalage B, et al. Évaluation d'une formation de proximité à la mesure de la clarté nucale et à l'autoévaluation des clichés. J Gynécologie Obstétrique Biol Reprod. 2010;39 (September 5):379–86.
[24] Alain D. Concordance auto évaluation et évaluation par un tiers pour les clichés biométriques d'échographie de dépistage du deuxième et troisième trimestre de la grossesse. [cited 2016 Jan 31]; Available from:. 2020. http://www.edu.upmc.fr/medecine/pedagogie/memoire/Memoires%2013/Memoire%20Dr_Daher.pdf.
[25] Gurusamy K, Aggarwal R, Palanivelu L, Davidson B. Virtual reality training for surgical trainees in laparoscopic surgery. In: The Cochrane Collaboration, editor. Cochrane database of systematic reviews [internet]. Chichester, UK: John Wiley & Sons, Ltd; 2007, doi:http://dx.doi.org/10.1002/14651858.CD006575 [cited 2018 Dec 2]. Available from:.
[26] Lentz GM, Mandel LS, Lee D, Gardella C, Melville J, Goff BA. Testing surgical skills of obstetric and gynecologic residents in a bench laboratory setting: validity and reliability. Am J Obstet Gynecol 2001;184(June 7):1462–70.
[27] Qureshi RN, Ali SK. Assessment of competence for caesarean section with global rating scale. J Pak Med Assoc 2013;63(8):6.
[28] Ofili-Yebovi D, Ben-Nagi J, Sawyer E, Yazbek J, Lee C, Gonzalez J, et al. Deficient lower-segment Cesarean section scars: prevalence and risk factors. Ultrasound Obstet Gynecol 2008;31(January 1):72–7.
[29] Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998;73(September 9):993–7.
[30] Gosman GG. Innovative approaches to resident education. Clin Obstet Gynecol 2003;46(March 1):15–30.
[31] El Haloui O, Delotte J, Gillard C, Boukaïdi S, Bongain A, Boucoiran I. [Instrumental extractions using Thierry's spatulas: evaluation of the learning curve]. Gynecol Obstet Fertil 2015;43(January 1):3–7.
[32] Gosman GG, Simhan HN, Guido RS, Lee TTM, Mansuria SM, Sanfilippo JS. Focused assessment of surgical performance: difficulty with faculty compliance. Am J Obstet Gynecol 2005;193(November 5):1811–6.
[33] Goff BA, Nielsen PE, Lentz GM, Chow GE, Chalmers RW, Fenner D, et al. Surgical skills assessment: a blinded examination of obstetrics and gynecology residents. Am J Obstet Gynecol 2002;186(April 4):613–7.
[34] van der Vleuten CPM, Schuwirth LWT, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. Best Pract Res Clin Obstet Gynaecol 2010;24(December 6):703–19.
[35] Bodle DJF, Kaufmann SJ, Bisson D, Nathanson B, Binney DM. Value and face validity of objective structured assessment of technical skills (OSATS) for work based assessment of surgical skills in obstetrics and gynaecology. Med Teach 2008;30(January 2):212–6.
[36] Chou B, Bowen CW, Handa VL. Evaluating the competency of gynecology residents in the operating room: validation of a new assessment tool. Am J Obstet Gynecol 2008;199(November 5)571 e1-571.e5.
[37] Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, et al. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10. Med Teach 2008;30(January 2):124–45.
[38] Adrales GL, Park AE, Chu UB, Witzke DB, Donnelly MB, Hoskins JD, A Valid Method of Laparoscopic Simulation Training and Competence Assessment 1, 2 1Presented at the 36th Annual Meeting of the Association for Academic Surgery. . p. 7–9 November 2Supported in part by an educational grant from Tyco/U.S. Surgical Corporation. J Surg Res. 2003 Oct;114(2):156–162.
[39] Maithel S, Sierra R, Korndorffer J, Neumann P, Dawson S, Callery M, et al. Construct and face validity of MIST-VR, Endotower, and CELTS: are we ready for skills assessment using simulators? Surg Endosc 2006;20(January 1):104–12.
[40] Mandel LS, Goff BA, Lentz GM. Self-assessment of resident surgical skills: is it feasible? Am J Obstet Gynecol 2005;193(November 5):1817–22.