Validation of a newly developed competency assessment tool for the Posterior Sagittal Anorectoplasty

Maja Joosten (maja.joosten@radboudumc.nl)
Radboudumc https://orcid.org/0000-0002-1674-1546

Guus MJ Bökkerink
Princess Maxima Center for Pediatric Oncology: Prinses Maxima Centrum voor Kinderoncologie

Jonathan Sutcliffe
Leeds Teaching Hospitals NHS Trust

Marc A Levitt
Children's National Medical Center: Children's National Health System

Karen A Diefenbach
Nationwide Children's Hospital

Carlos A Reck
Medical University of Vienna: Medizinische Universitat Wien

Wilfried Krois
Medical University of Vienna: Medizinische Universitat Wien

Ivo de Blaauw
Radboudumc

Sanne MBI Botden
Radboudumc

Research article

Keywords: assessment, pediatric surgery, posterior sagittal anoplasty

Posted Date: October 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-97670/v1

Abstract

Background: The correction of an anorectal malformation (ARM) is highly complex and relatively infrequent. Simulation training and subsequent assessment of its effectiveness may result in better clinical outcomes. Assessment can be done using a CAT to assess surgical skills of a specific procedure. This study aims to develop and validate a competency assessment tool (CAT) for the posterior sagittal anorectoplasty (PSARP) for the repair of anorectal malformation.

Methods: The CAT-PSARP was developed after consultation with experts in the field. The PSARP was divided into five steps, while general skills (tissue and instrument handling) were scored separately. During several pediatric colorectal hands-on courses in 2019 and 2020 participants were asked to voluntarily participate. They performed one PSARP procedure on an ARM simulation model, while being assessed by two objective observers using the CAT-PSARP.

Results: A total of 82 participants were enrolled. Intra-class correlation coefficient analysis demonstrated a fair inter-observer agreement for 'total score of general skills' (ICC =0.524, $p<0.001$), a good agreement for 'total score of specific skills' (ICC=0.646 , $p<0.001$) and 'total score overall' (ICC=0.669, $p<0.001$). When comparing the experienced with the inexperienced group a significantly higher score for the experienced group was found for all steps ($p=0.013$ to $p<0.001$), except for step 5 'anoplasty' ($p=0.440$).

Conclusion: Good inter-observer agreement was found for the total scores of the CAT-PSARP. It is a suitable objective assessment tool for the overall performance of the included steps of the PSARP for repair of an ARM.

Background

In surgical training programs, the development of surgical skills in the operating theatre plays a pivotal role [1]. In order to adequately assess the development of these skills, the format of surgical training is structured as a competency-based curricula. This results in a strong need for educational tools that provide objective assessment of surgical performance to evaluate predefined competency goals [2–6]. Over the past decades, various assessment methods have been developed to fulfil this goal, including tracking devices, virtual reality simulators, observational instruments, and computer games [7–10]. Currently, skills of surgeons in training are often assessed in the clinical setting by experts using the Objective Structured Assessment of Technical Skills (OSATS) form, based on the overall performance [11–13]. However, tools like OSATS and its derivatives are not specifically designed to provide information on the separate skills being trained. Subsequently, OSATS do not provide any formative information or feedback on the skills that still need improvement or the skills that are already sufficient. This results in unawareness of trainees regarding which specific skills need improvement. Furthermore, there is no clear correlation between the OSATS score and the outcome of the specific procedure that the resident or surgeon has performed [14]. Another assessment tool that can be used to assess the surgical skills is a competency assessment tool (CAT). This is a method of assessing the surgical performance by

describing specific steps of a given procedure. It evaluates both the process of performance (instrument use, tissue handling, and committed errors) and the quality of the end product. The CAT has been successfully applied to improve the quality of training in the English National Training Program for laparoscopic colorectal surgery [15]. It has also been further developed and used for the assessment of the laparoscopic cholecystectomy [13, 16] and for laparoscopic suturing [17].

There are currently no published assessment tools that specifically assess the performance of pediatric surgical procedures. One procedure in which technical proficiency correlates with clinical outcome is the Posterior Sagittal Anorectoplasty (PSARP) in infants with an anorectal malformation. Due to the rareness of this congenital malformation, surgical trainees have limited exposure to the complex surgical procedure. The aim of this study is to develop and validate a new competency assessment tool for the objective assessment of the completion of a PSARP for the correction of an anorectal malformation.

Methods

Development of the CAT-PSARP

The CAT-PSARP was developed with a structured approach, based on the previously developed CAT's [13, 15–17]. The focus of this CAT-PSARP form is on the performance of each step as well as the general skills separately.

Multiple experts in the field of pediatric colorectal surgery were consulted to assess the steps used in the form and the separate assessment items. The five steps were standardised and agreed upon, before subdividing these in gradings. The grading system was discussed with the experts, to reach consensus on the form. The PSARP was divided into the following five component steps which were incorporated in the CAT-PSARP as the task specific skills:

Step 1: Sagittal opening in midline

Step 2: Placing sutures around fistula

Step 3: Dissection fistula/ rectum

Step 4: Reconstruction of sphincter complex

Step 5: Anoplasty

Based on the previous CAT's a score of the general skills was also included on instrument and tissue handling, although these were not weighed as heavily compared to the task specific skills. The general opinion was that the task specific skills and incorporation of a structured approach to the PSARP procedure was more important in correction of an anorectal malformation, than perfect instrument movement and tissue handling alone. In this CAT-PSARP the instrument handling and tissue handling are scored as an overall score, therefore these are doubled (2-4-6-8) compared to the scores of each

component step (1-2-3-4). The higher the score, the better the performance. The previous CAT's had a lower score for the best performances, but the opinion of the experts was that this was confusing and all preferred a higher score for a better performance. Figure 1 shows the final assessment tool and grading system.

ARM simulation model

The ARM simulation model, with perineal fistula, used in this study was a validated partly reusable low-budget model, which consisted of a wooden laser cut casing, with a disposable perineal body (PediatrickBoxx, Fig. 2). The casing is reusable and of light weight material. The perineal body is a single-use, replaceable model consisting of layered sponges and silicone, with a double layered balloon to represent the recto-perineal fistula. Each participant received their own perineal body to complete the procedure.

Participants

The participants of the hands-on workshops during the 'Regional pediatric surgery course', Nijmegen, the Netherlands, October 2019, 'Pediatric Colorectal and Pelvic Reconstruction Congress', Columbus, Ohio, November 2019, the '12th European Pediatric Colorectal Congress', Vienna, December 2019 and the 'Pediatric Colorectal Course' Nijmegen, March 2020 were asked to participate in this study. Participation in this study was voluntarily and independent of the supervision and training during the workshop. According to local law and legislation ethical board approval was waived, written informed consent was obtained from all participants. It was explained that the assessment score was for research purposes only, was not used during the course, and was processed anonymously.

The participants were divided into groups based on their experience, used for further subanalyses. Participants were included in the experienced group if they had performed ≥ 20 colorectal reconstructions and ≥ 5 PSARP procedures. Participants were included in the inexperienced group if they had performed ≤ 5 PSARP procedures and had performed ≤ 5 colorectal reconstructions. The remainder of the participants were in the intermediate group, which did have a background in pediatric surgery, but were not experienced in pediatric colorectal surgery.

Protocol

The participants of the study all completed a short questionnaire on their demographics and previous clinical experience, particularly their pediatric colorectal experience. All participants were shown an instruction video of the steps of the PSARP that should be performed on the model. Additionally, a poster was developed of the steps to be assessed, to guide the trainees during the training on the model. This was also printed on the back side of the CAT-PSARP form as an aid for the assessor. The participants received a short introduction on the CAT-PSARP form, to acquaint them with the demands of each step of the procedure.

All participants performed one PSARP procedure on the model, while two independent objective observers assessed the component steps during their performance on the CAT-PSARP. The observers were selected based on availability from a pool of five experts, which alternated to keep bias as low as possible. The CAT-PSARP forms were coded with the participant and observer number, to be able to connect these to the demographics, but were anonymized for the researchers. The demographic data of the participants was not available for the experts before, during, or after the assessment to avoid bias in the scoring of the performance. All data was processed anonymously and all participants provided approval for inclusion in the study.

Statistical analysis

All statistical analyses were performed with IBM's SPSS statistics v.25 package. The inter-observer agreement for the component steps as well as for the general and total scores between the two observers was assessed using the intra-class correlation coefficient (ICC), with a two-way Random Effects model, on a two-tailed significance level of $p < 0.05$. An ICC value of < 0.4 was considered a poor agreement, ≥ 0.4 and < 0.6 was considered a fair agreement, ≥ 0.6 and < 0.8 a good agreement and ≥ 0.8 an excellent agreement.

For construct validity the scores of inexperienced, intermediate and experienced participants were compared with a one-way ANOVA. Equality of variances was assumed if Levene's test for equality of variance was > 0.05. This process was conducted by an independent researcher who was not involved in the scoring process using the completed CAT-PSARP forms of the objective observers.

The aim was to include at least 30 participants for the interobserver reliability [18].

Results

A total of 82 participants completed the PSARP on the PediaTrickBoxx ARM model and were scored independently by two objective observers (observer A and B). The majority (48%) were pediatric surgeons, 33% were surgical residents and 15% were fellows pediatric surgery. Male-female ratio was equally divided between the participants, with a mean age of 36 years, and the majority (76%) was European, including 29% Dutch and 18% German participants (Table 1). Of the total group, 41% had never performed a colorectal reconstruction and 52% had never performed a PSARP procedure. In contrast to that, five participants had performed > 50 colorectal reconstructions and two participants performed > 50 PSARP procedures. The detailed pediatric colorectal experience is shown in Table 2.

Table 1
Demographic properties of the study population, variables are stated as number (percentage) or mean (standard deviation). * 1 missing.

Demographic properties	Total group n = 82	Inexperienced n = 22	Intermediate n = 44	Experienced n = 16
Age in years, mean (SD)	36.4 (6.0)	33.9 (6.2)	37.0 (5.3)	38.5 (6.9)
Gender, N (%)				
Female	40 (49)	10 (46)	25 (57)	5 (31)
Male	42 (51)	12 (54)	19 (43)	11 (69)
Profession, N (%)*				
Pediatric surgeon	39 (48)	4 (18)	22 (51)	13 (81)
Pediatric urologist	2 (2.5)	0 (0)	0 (0)	2 (13)
General surgeon	1 (1.2)	0 (0	1 (2.3)	0 (0)
Fellow pediatric Surgery	12 (15)	1 (4.5)	10 (23)	1 (6.7)
Surgical resident	27 (33)	17 (77)	10 (23)	0 (0)
Years in this profession, mean (SD)	5.0 (4.3)	2.6 (6.2)	5.6 (4.9)	5.6 (4.1)
Place of residence, N (%)*				
Europe	64 (78)	20 (91)	34 (79)	10 (63)
North-America	9 (11)	1 (4.5)	6 (14)	2 (13)
Africa	4 (4.9)	0 (0)	2 (4.6)	2 (13)
Asia	3 (3.7)	1 (4.5)	1 (2.3)	1 (6.3)
Australia	1 (1.2)	0 (0)	0 (0)	1 (6.3)

Table 2
Pediatric colorectal experience of the participants.

Pediatric colorectal experience	Total group n = 82	Inexperienced n = 22	Intermediate n = 44	Experienced n = 16
Colorectal reconstructions performed, N (%)				
None	32 (41)	16 (73)	16 (36)	0 (0)
< 5	18 (23)	6 (27)	12 (27)	0 (0)
5–20	16 (20)	0 (0)	16 (36)	0 (0)
20–50	11 (13)	0 (0)	0 (0)	11 (69)
> 50	5 (6)	0 (0)	0 (0)	5 (31)
Colorectal reconstructions assisted, N (%)*				
None	23 (30)	14 (64)	8 (20)	1 (6)
< 5	14 (18)	5 (23)	9 (23)	0 (0)
5–20	22 (28)	2 (9)	15 (38)	5 (31)
20–50	12 (15)	1 (5)	6 (15)	5 (31)
> 50	7 (9)	0 (0)	2 (5)	5 (31)
PSARP performed, N (%)**				
None	41 (52)	19 (86)	22 (54)	0 (0)
< 5	17 (22)	3 (14)	14 (34)	0 (0)
5–20	11 (14)	0 (0)	5 (12)	6 (38)
20–50	8 (10)	0 (0)	0 (0)	8 (50)
> 50	2 (3)	0 (0)	0 (0)	2 (13)
PSARP assisted, N (%)*				
None	25 (32)	16 (73)	9 (23)	0 (0)

* 4 missing, ** 3 missing.

Inexperienced group: ≤5 colorectal reconstructions performed, ≤ 5 PSARP performed

Experienced group: ≥20 colorectal reconstructions performed, ≥ 5 PSARP performed

Intermediate group: all other participants

Pediatric colorectal experience	Total group n = 82	Inexperienced n = 22	Intermediate n = 44	Experienced n = 16
< 5	22 (28)	5 (23)	15 (38)	2 (13)
5–20	20 (26)	1 (5)	14 (35)	5 (31)
20–50	7 (9)	0 (0)	2 (5)	5 (31)
> 50	4 (5)	0 (0)	0 (0)	4 (25)
PSARP on animals, N (%)*				
None	69 (89)	21 (95)	34 (83)	14 (88)
< 5	8 (10)	1 (5)	6 (15)	1 (17)
5–20	1 (1)	0 (0)	1 (2)	1 (17)
20–50	0 (0)	0 (0)	0 (0)	0 (0)
> 50	0 (0)	0 (0)	0 (0)	0 (0)
* 4 missing, ** 3 missing.				
Inexperienced group: ≤5 colorectal reconstructions performed, ≤ 5 PSARP performed				
Experienced group: ≥20 colorectal reconstructions performed, ≥ 5 PSARP performed				
Intermediate group: all other participants				

Inter-observer reliability

When looking at the outcomes of the component steps separately (Table 3), the intra-class correlation coefficient revealed a poor inter-observer agreement for ‘opening in the midline’ (ICC = 0.289) and ‘dissection of the fistula’ (ICC = 0.148). A fair inter-observer agreement was found for ‘placing traction sutures on the fistula’ (ICC = 0.557). A good agreement was found for the steps ‘closure of the perineal body and sphincter complex’ (ICC = 0.602) and ‘anoplasty’ (ICC = 0.605). Combining the results of the component steps resulted in a good the inter-observer agreement with an ICC of 0.646 ($p < 0.001$).

Table 3
Consistency (intra-observer and inter-observer variability) between the scores given by expert A and expert B, depicted as mean with standard deviation (SD) and standard error of the mean (SEM), calculated with the intraclass correlate coefficient (ICC), on a 2-tailed significance level of $p < 0.05$.

	Expert A n = 82	Expert B n = 82	ICC	p-value
Total score general skills[1]	11 (2.5; 0.3)	11 (2.2; 0.2)	0.524	< 0.001
Instrument handling	5.3 (1.3; 0.1)	5.7 (1.1; 0.1)	0.475	< 0.001
Tissue handling	5.5 (1.4; 0.2)	5.6 (1.2; 0.1)	0.574	< 0.001
Total score specific skills[2]	15 (2.4; 0.3)	14 (1.8; 0.2)	0.646	< 0.001
Opening in midline	3.3 (0.5; 0.1)	3.1 (0.6; 0.1)	0.289	0.005
Traction sutures on fistula	2.6 (1.0; 0.1)	2.6 (0.9; 0.1)	0.557	< 0.001
Dissection of fistula	3.2 (0.5; 0.1)	3.2 (0.4; 0.0)	0.148	0.099
Closure of perineal body and sphincter complex	3.1 (0.7; 0.1)	2.8 (0.6; 0.1)	0.602	< 0.001
Anoplasty	2.9 (0.9; 0.1)	2.7 (0.8; 0.1)	0.605	< 0.001
Total score overall[3]	26 (3.7; 0.4)	26 (3.4; 0.4)	0.669	< 0.001

[1] General skills: sum of the tissue handling and instrument handling

[2] Specific skills: 5 steps of the PSARP

[3] Overall: sum of general skills and specific skills.

When focussing on the general skills rated on the CAT-PSARP, these showed a fair inter-observer agreement (instrument handling ICC = 0.475 tissue handling ICC = 0.574), with also a fair inter-observer agreement for the 'total score of general skills' (ICC = 0.523, $p < 0.001$). The evaluation of the CAT-PSARP is aimed at the total score of all items on the form, which demonstrated a good inter-observer agreement with an ICC of 0.669 ($p < 0.001$), as shown in Table 3.

Construct validity

To evaluate the construct validity of this assessment form, the results of the three expertise groups (inexperienced, intermediate, experienced) were compared. The scores of the three groups were significantly different for all items ($p = 0.038$ to $p < 0.001$) as shown in Table 4. A subsequent post-hoc analysis was performed to evaluate these differences. As seen in Table 5 and Fig. 3, the total score, the overall score of the component steps and the general skills were significantly different between all three groups. The experienced group scored significantly higher than the inexperienced group for all separate items except for step 5 'anoplasty' ($p = 0.054$). The experienced group also scored significantly higher

than the intermediate group, who had some pediatric colorectal experience but still did not perform as well as the experienced participants, with significant differences in the overall scores, but also in the separate general skills ($p < 0.001$) and the component steps 'dissection of the fistula' ($p < 0.001$) and 'anoplasty' ($p = 0.038$). The inexperienced group scored significantly lower than the intermediate group for 'opening in midline' ($p = 0.022$), placing traction sutures on fistula' ($p < 0.001$) and 'closure of the perineum' ($p = 0.017$).

Table 4
Scores of inexperienced, intermediate and experienced group to determine construct validity, depicted as mean with standard deviation (SD), differences calculated with a one-way ANOVA, on a 2-tailed significance level of $p < 0.05$.

	Inexperienced **N = 21**	**Intermediate** **N = 44**	**Experienced** **N = 16**	**p-value**
Total score general skills[1]	10.2 (1.9)	10.7 (1.6)	13.2 (1.7)	< 0.001
Instrument handling	5.2 (0.2)	5.3 (0.8)	6.3 (0.9)	< 0.001
Tissue handling	5.0 (1.2)	5.4 (0.9)	6.8 (0.9)	< 0.001
Total score specific skills[2]	13.2 (2.0)	14.6 (1.2)	16.4 (1.9)	< 0.001
Opening in midline	3.0 (0.6)	3.3 (0.3)	3.4 (0.4)	0.001
Traction sutures on fistula	2.0 (0.8)	2.8 (0.7)	3.1 (0.8)	< 0.001
Dissection of fistula	3.1 (0.4)	3.1 (0.2)	3.5 (0.4)	< 0.001
Closure of perineal body and sphincter complex	2.5 (0.7)	2.9 (0.6)	3.3 (0.5)	< 0.001
Anoplasty	2.9 (0.9)	2.6 (0.7)	3.1 (0.8)	0.038
Total score overall[3]	23.5 (3.1)	25.4 (1.8)	29.7 (2.7)	< 0.001

[1] General skills: sum of the tissue handling and instrument handling

[2] Specific skills: 5 steps of the PSARP

[3] Overall: sum of general skills and specific skills.

Table 5
P-values of post-hoc analysis using Hochberg's GT2 on a 2-tailed significance level of $p < 0.05$. Inexperienced n = 21; intermediate n = 44; experienced n = 16.

	Inexperienced vs. intermediate	Intermediate vs. experienced	Inexperienced vs. experienced
Total score general skills[1]	0.625	< 0.001	< 0.001
Instrument handling	0.996	< 0.001	0.001
Tissue handling	0.292	< 0.001	< 0.001
Total score specific skills[2]	0.001	< 0.001	< 0.001
Opening in midline	0.022	0.270	0.001
Traction sutures on fistula	< 0.001	0.115	< 0.001
Dissection of fistula	0.964	< 0.001	0.001
Closure of perineal body and sphincter complex	0.017	0.082	< 0.001
Anoplasty	0.486	0.038	0.540
Total score overall[3]	0.003	< 0.001	< 0.001

[1] General skills: sum of the tissue handling and instrument handling

[2] Specific skills: 5 steps of the PSARP

[3] Overall: sum of general skills and specific skills.

Discussion

This CAT-PSARP has shown to be a suitable scoring tool for the PSARP procedure in a simulation setting, with a good overall inter-observer reliability. The majority of component steps showed a good inter-observer agreement, as well as the general performance skills. By comparing the experienced with the inexperienced participants a construct validity of this assessment tool was also established, because significant differences were found, which were clinically relevant. The differences in the performance between experienced and inexperienced participants were three points on both the specific and the general skills, with a total difference of six points on average. This indicates a possibility for training towards a score goal of 30 (out of a possible total of 36). By adding the intermediate group, which were the real target group for assessments, we showed that small differences in performances could also be assessed with this CAT-PSARP form. There were significant differences between all three expertise groups in total scores of the form, which indicates that this is a potent tool for the assessment of pediatric surgery trainees throughout their training.

Although a poor inter-observer agreement was found for the steps ‘opening in midline’ and ‘dissection of the fistula’, this could be related to the difficulty in simulating these steps on this low-budget, PediaTrickBoxx ARM simulation model. The vast majority of the participants scored the maximum amount of four points on both of these component steps due to the fact that they were fairly easy in comparison with the real procedure. In the clinical setting, the distribution of the scores for these steps will likely differ more amongst the trainees, which may result in an improvement of the intra-observer correlation [19].

Overall, this CAT-PSARP had slightly lower inter-observer reliability scores compared to the other previously developed CAT’s [13, 15–17]. There are several possible explanations for this. First, this is a CAT for open surgery and assessments are done in real time. In contrast, when assessing minimally invasive surgery (MIS), it is possible to use a video of the performance for the assessment, which results in the possibility for a more meticulous assessment by pausing or replaying the video. In addition, this newly developed CAT-PSARP was used for assessment of all participants of the course at the same time, up to a maximum of ten participants simultaneously. If the CAT-PSARP will be used in the clinical setting, this could be an advantage, because the supervisor (who has seen the whole procedure) can complete the CAT-PSARP if the trainee has done (part of) the procedure. Further research will evaluate whether implementation of the CAT-PSARP in clinical setting is feasible and whether it ultimately results in an improved performance of pediatric surgical trainees over time. Second, based on the consensus of pediatric colorectal experts, this CAT-PSARP focuses on the component steps of the PSARP and tissue and instrument handling are scored separately. This is in contrast to previous CAT’s, in which each step was only scored on tissue and instrument handling. This is because it is believed that the task specific skills and incorporation of a structured approach in the PSARP procedure was more important in performing a technically proficient PSARP, than perfect instrument and tissue handling alone. However, general skills were taken into account for the complete score of overall performance. It has previously been suggested that procedure specific assessments are more useful for trainees than a general assessment. Kamran et al. stated that the combination of global skills (e.g. tissue and instrument handling) as well as task specific skills (such as the steps of the PSARP) may provide a more comprehensive and concise feedback to the trainee than an general scoring system, only evaluating general skills [7].

Limitations

There are some limitations to this study. The assessment was done while participants practice the PSARP procedure on a simulation model. It should be researched whether it can be used in clinical setting as well. In addition, the component steps were defined after discussion with experts in the field. However, there may be other aspects of the repair of an ARM with another type of approach (such as an anterior-sagittal anorectoplasty (ASARP)) that were not addressed in this CAT. The two independent objective observers were selected from a group of five objective observers to minimize risk of bias, however, this could not be completely eliminated. All received the same brief explanation on how to use the CAT-PSARP

shortly before starting the assessment. This design limits the risk of bias and is true to the clinical setting where trainees are assessed by different supervisors as well [20].

Conclusion

The CAT-PSARP showed a good overall inter-observer agreement as well as a good construct validity, which makes it a potent assessment tool. The combination of assessment of both specific component tasks of the procedure and general skills, including tissue and instrument handling, showed to be most relevant in the assessment of the repair of an anorectal malformation by posterior sagittal approach.

Abbreviations

ARM anorectal malformation

ASARP anterior-sagittal anorectoplasty

CAT competency assessment tool

ICC intra-class correlation coefficient

MIS minimally invasive surgery

OSATS Objective Structured Assessment of Technical Skills

PSARP posterior sagittal anorectoplasty

SEM standard error of the mean

SD standard deviation

Declarations

Ethics approval and consent to participate

Written informed consent was obtained from all participants. According to national law and legislation ethical board approval of the ethics committee of Arnhem and Nijmegen was deemed unnecessary Approval of the ethics committee of the institution Radboudumc was waived according to national regulations [21].

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

Maja Joosten, Guus MJ Bökkerink, Jonathan Sutcliffe, Marc A Levitt, Karen A Diefenbach, Carlos A Reck, Wilfried Krois, Ivo de Blaauw, Sanne MBI Botden declare that they have no competing interests.

Funding

Maja Joosten, Guus MJ Bökkerink, Jonathan Sutcliffe, Marc A Levitt, Karen A Diefenbach, Carlos A Reck, Wilfried Krois, Ivo de Blaauw, Sanne MBI Botden have no financial ties to disclosure. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' contribution

Study concept and design: MJ, IB, SB. Data acquisition: MJ, GB, JS, ML, KD, CR, WK, IB, SB.

Analysis and data interpretation: MJ. Drafting of the manuscript: MJ

Critical revision: GB, JS, MA, KD, CR, WK, IB, SB

All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Baldwin P, Paisley A, Brown S. Consultant surgeons' opinion of the skills required of basic surgical trainees. Br J Surg. 1999;86:1078–82.
2. Brennan M, Debas H. Surgical education in the United States: portents for change. Ann Surg. 2004;240:565–72.
3. Collins J, Civil I, Sugrue M, et al. Surgical Education and Training in Australia and New Zealand. World J Surg. 2008;32(10):2138–44.
4. Frank J, Langer B. Collaboration, communication, management, and advocacy: teaching surgeons new skills through the CanMEDS Project. World J Surg, 2003;972–978.
5. Smith A, Aggarwal R, Warren O, et al. Surgical Training and Certification in the United Kingdom. World J Surg. 2009;33(2):174–9.
6. Epstein R. Assessment in medical education. N Engl J med. 2007;356:387–96.
7. Ahmed K, Miskovic D, Darzi A, et al. Observational tools for assessment of procedural skills: a systematic review. Am J Surg. 2011;202:469–806.

8. Datta V, Mackay S, Mandalia M, et al. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg. 2001;193:479–85.
9. Aggarwal R, Grantcharov T, Eriksen J, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg. 2006;244:310–4.
10. Graafland M, Schraagen J, Schijven M. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012;99:1322–30.
11. Eriksen J, Grantcharov T. Objective assessment of laparoscopic skills using a virtual reality stimulator. Surg Endosc. 2005;19:1216–9.
12. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. Br J Surg. 2004;91:1549–58.
13. Ganni S, Chmarra M, Goossens R, et al. Self-assessment in laparoscopic surgical skills training: Is it reliable? Surg Endosc. 2017;31(6):2451–6.
14. Anderson D, Long S, Thomas G, et al. Objective Structured Assessments of Technical Skills (OSATS) Does not assess the quality of the surgical result effectively. Clin Orthop Relat Res. 2016;474(4):874–81.
15. Miskovic D, Ni M, Wyles S, et al. Is competency assessment at the specialist level achievable? A study for the national training programme in laparoscopic colorectal surgery in England. Ann Surg. 2013;257(3):476–82.
16. Ganni S, Botden S, Schaap D, et al. Improves Self-Assessment and End-Performance in Laparoscopic Surgical Skills Training. J Surg Educ. 2018;75(2):527–33.
17. IJgosse WM, Leijte E, Ganni S, et al. Competency assessment tool for laparoscopic suturing: development and reliability evaluation. Surg Endosc, 2019;1–7.
18. Koo T, Li M. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine. 2016;15(2):155–63.
19. Popović Z, Thomas J. Assessing observer variability: a user's guide. Cardiovascular diagnosis therapy. 2017;7(3):317–24.
20. Mahtani K, Spencer E, Brassey J, et al. Catalogue of bias: observer bias. BMJ. 2018;23(1):23.
21. WMO. https://wetten.overheid.nl/BWBR0009408/2020-01-01 accessed on 09-10-2020.

Figures

General skills			Score	Task specific skills	Score		
Form number: Number performer: _____ Number sponge: _____ Observer: Self – Peer – Expert Number observer: _____ Signature: _____				**1. Opening in midline**	1	Off center	From start off center
					2	Off center	Starting in midline, but subcutaneously off center
					3	Little off center	Corrected during procedure
					4	Perfect in midline	Whole procedure in midline
				2. Traction sutures on fistula	1	Wrong placement on edge	Too large bites of surrounding skin
					2	Wrong placement on edge	Too little bite on fistula
					3	Inconsistent bites	But not too little or too big
					4	Proper in plane bites	Circumferentially of the fistula
Instrument handling	2	Uncoordinated movements	Stiff and uncontrolled	**3. Dissection fistula**	1	Multiple perforations	>1x in wrong plane on rectum/ fistula
	4	Hesitant movements	Controlled, but ineffective		2	1 perforation	Once in wrong plane on rectum/ fistula
	6	Skillful movements	Smooth, controlled and meaningful		3	Proper dissection plane	Eventually in proper dissection plane
	8	Versatile movements	Masterful and effective instrument use		4	Proper dissection plane	During whole procedure
Tissue handling	2	Stagnant	No guidance with non-dominant hand, unstructured closure	**4. Closure of perineal body and sphincter complex**	1	No closure	No sutures to close perineum and sphincter
	4	Lagging	Adjusting with delay, no counter traction		2	Mal-aligned closure	Few non structured sutures in subcutis or no rectum tagging
	6	Meaningful	Keeps improving the exposure		3	Closure	Using the triangle technique non-structured with rectum tagging
	8	Forward looking	Strategic and intelligent adjustments, structured and precise closure		4	Structured closure	Using the triangle technique structured with rectum tagging
Score general skills				**5. Recto-cutaneous anastomosis**	1	Mal-aligned anastomosis	Anastomosis edges not aligned, large gaps
Score specific skills					2	Poor anastomosis	<16 or >16 sutures, not equally divided or inadequate sized anastomosis
Total score					3	Adequate anastomosis	16 sutures, not equally divided, adequate size anastomosis
					4	Perfect anastomosis	16 sutures, equally divided, no gaps, adequate size anastomosis

Figure 1

CAT-PSARP from for assessing the PSARP. The form is divided in ‘general skills’ (instrument handling and tissue handling) which can be scored 2 to 8 and ‘specific skills’ (the five component steps of the PSARP) which can be scored 1 to 4.

Figure 2

PediatrickBoxx ARM model for training of the PSARP. The model consists of a wooden casing with a silicone perineal body.

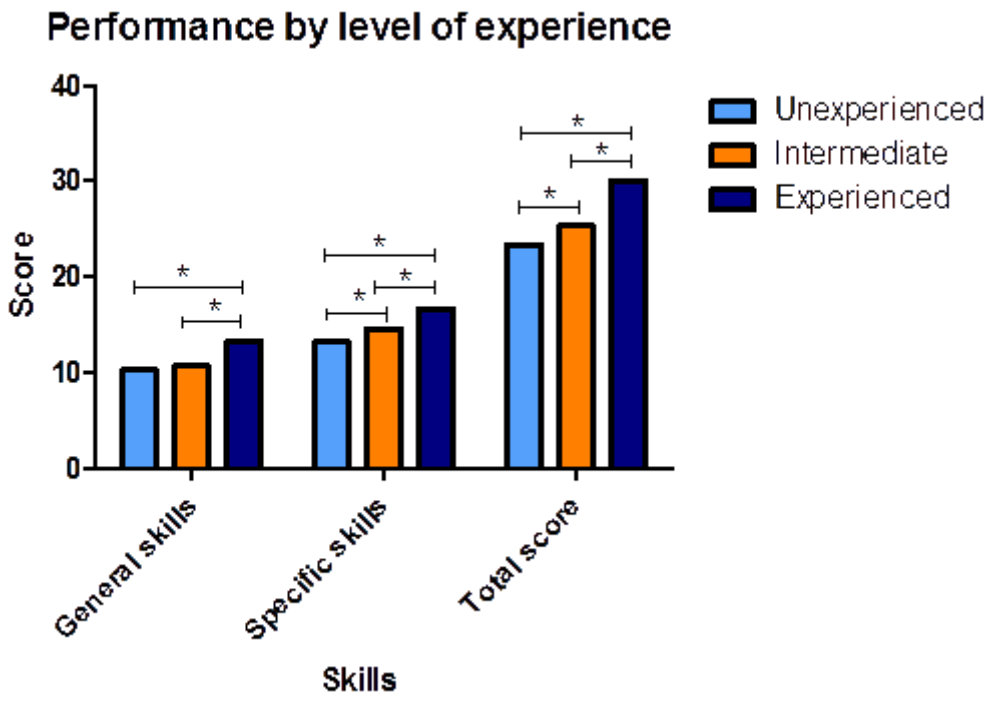

Figure 3

Mean scores of inexperienced, intermediate and experienced participants, for the general skills, specific skills and total score. Asterisk indicates a significant difference in mean scores between two groups.