Validation of a Newly Developed Competency Assessment Tool for the Posterior Sagittal Anorectoplasty

Maja Joosten1 Guus M.J. Bökkerink2 Jonathan Sutcliffe3 Marc A. Levitt4 Karen Diefenbach5 Carlos A. Reck6 Wilfried Krois6 Ivo de Blaauw7 Sanne M.B.I. Botden8

1 Department of Pediatric Surgery, Radboud University Medical Center, Nijmegen, The Netherlands
2 Department of Pediatric Surgery, Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands
3 Department of Paediatric Surgery, Leeds Children’s Hospital, F Floor, Martin Wing Leeds General Infirmary GT George Street, Leeds, W Yorkshire, United Kingdom
4 Department of Surgery, Colorectal and Pelvic Reconstructive Surgery, Children’s National Hospital, District of Columbia, Washington, United States
5 Department of Pediatric Surgery, Nationwide Children’s Hospital, Columbus, Ohio, United States

Address for correspondence Maja Joosten, MD, Department of Pediatric Surgery, Radboud University Medical Center, Geert Grooteplein Zuid 10 route 618, 6500HB Nijmegen, The Netherlands (e-mail: maja.joosten@radboudmc.nl).

6 Department of Pediatric Surgery, Medical University of Vienna, Wien, Wien, Austria
7 Department of Pediatric Surgery, Radboud Medical Centre, Nijmegen, The Netherlands
8 Department of Pediatric Surgery, Radboudumc-Amalia Children’s Hospital, Nijmegen, The Netherlands

Eur J Pediatr Surg

Abstract

Introduction The correction of an anorectal malformation (ARM) is complex and relatively infrequent. Simulation training and subsequent assessment may result in better clinical outcomes. Assessment can be done using a competency assessment tool (CAT). This study aims to develop and validate a CAT for the posterior sagittal anorectoplasty (PSARP) on a simulation model.

Materials and Methods The CAT-PSARP was developed after consultation with experts in the field. The PSARP was divided into five steps, while tissue and instrument handling were scored separately. Participants of pediatric colorectal hands-on courses in 2019 and 2020 were asked to 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. A fair interobserver agreement was found for general skills (intraclass correlation coefficient [ICC] = 0.524, \( p < 0.001 \)), a good agreement for specific skills (ICC = 0.646, \( p < 0.001 \)), and overall performance (ICC = 0.669, \( p < 0.001 \)). The experienced group scored higher on all steps (\( p < 0.001 \)), except for “anoplasty (\( p = 0.540 \)),” compared with an inexperienced group.

Conclusion The CAT-PSARP is a suitable objective assessment tool for the overall performance of the included steps of the PSARP for repair of an ARM on a simulation model.

Keywords ► assessment ► pediatric surgery ► posterior sagittal anoplasty ► simulation model ► training
Introduction

In surgical training programs, the development of surgical skills in the operating theater plays a pivotal role.1 To adequately assess the development of these skills, the format of surgical training is structured as a competency-based curriculum. 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 fulfill 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 that are trained and no clear correlation between the OSATS score and the outcome of the specific procedure that the resident or surgeon has performed.14 The competency assessment tool (CAT) has advantages over this method, including description of specific steps of a given procedure, evaluating both performance (e.g., tissue handling), as well as 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 cholecystectomy13,16 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 (ARM). Due to the rareness of this congenital malformation, surgical trainees have limited exposure to this complex surgical procedure.18,19 The aim of this study is to develop and validate a new CAT for the objective assessment of the completion of a PSARP for the correction of an ARM.

Methods

Development of the Competency Assessment Tool–Posterior Sagittal Anorectoplasty

The CAT-PSARP was developed with a structured approach, based on the previously developed CATs.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 standardized and agreed on, 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.

The general opinion was that the task-specific skills and incorporation of a structured approach to the PSARP procedure were important in correction of an ARM. A score on the general skills (instrument handling and tissue handling) during the entire procedure was included as well. Because the general skills are scored as an overall score, these are doubled (2–4 to 6–8) compared with the scores of each component step (1–2 to 3–4) to assure equal weight of general skills and specific skills. The higher the score, the better the performance. The previous CATs 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. – Fig. 1 shows the final assessment tool and grading system.

Anorectal Malformation 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).20,21 The casing is reusable and made of light weight material. The perineal body is a single use and replaceable model consisting of layered sponges and silicone with a double layered balloon to represent the rectoperineal 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 in their surgical career. Participants were included in the inexperienced group if they had performed ≤ 5 PSARP procedures and had performed ≤ 5 colorectal reconstructions in their surgical career. 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 instructional 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 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 were not available for the experts before, during, or after the assessment to avoid bias in the scoring of the performance. All data were 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 interobserver agreements for the component steps, as well as for the general and total scores between the two observers, were assessed using the intraclass 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, \( \geq 0.4 \) and \(< 0.6 \) were considered a fair agreement, \( \geq 0.6 \) and \(< 0.8 \) a good agreement, and \( \geq 0.8 \) an excellent agreement.\(^{22}\)

For construct validity, the scores of inexperienced, intermediate, and experienced participants were compared with a one-way analysis of variance (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.\(^{23}\)

### Results

A total of 82 participants completed the PSARP on the PediaTrickBoxx ARM model and were scored independently by two objective observers (observers A and B). The majority (48%) were pediatric surgeons, 33% were surgical residents, and 15% were fellows of pediatric surgery. Male-to-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.

#### Interobserver Reliability

When looking at the outcomes of the component steps separately (Table 3), the ICC revealed a poor interobserver agreement for "opening in the midline (ICC = 0.289)" and "dissection of the fistula (ICC = 0.148)." A fair interobserver 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 interobserver agreement with an ICC of 0.646 (\( p < 0.001 \)).

When focusing on the general skills rated on the CAT-PSARP, these showed a fair interobserver agreement (instrument handling, ICC = 0.475 and tissue handling, ICC = 0.574), with also a fair interobserver 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 interobserver agreement with an ICC of 0.669 (\( p < 0.001 \)) as shown in Table 3.
To evaluate the construct validity of this assessment form, the results of the three expertise groups (inexperienced, intermediate, and 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, overall score of the component steps, and 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 component steps “dissection of the fistula (p < 0.001)” and “anoplasty.”

### Table 2 Pediatric colorectal experience of the participants

| Pediatric colorectal experience | Total group (n = 82) n (%) | Inexperienced (n = 22) n (%) | Intermediate (n = 44) n (%) | Experienced (n = 16) n (%) |
|--------------------------------|---------------------------|-----------------------------|----------------------------|----------------------------|
| Colorectal reconstructions performed |                           |                             |                            |                            |
| 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 assisteda |                         |                             |                            |                            |
| 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 performedb               |                           |                             |                            |                            |
| 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 assisteda               |                           |                             |                            |                            |
| None                          | 25 (32)                   | 16 (73)                     | 9 (23)                     | 0 (0)                      |
| < 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 animalsa              |                           |                             |                            |                            |
| 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)                      |

Abbreviation: PSARP, posterior sagittal anorectoplasty.
Notes: Inexperienced group: ≤ 5 colorectal reconstructions performed, ≤ 5 PSARP performed.
Experienced group: ≥ 20 colorectal reconstructions performed, ≥ 5 PSARP performed.
Intermediate group: all other participants.

a4 missing.
b3 missing.
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).”

**Discussion**

This CAT-PSARP has shown to be a suitable scoring tool for the PSARP procedure in a simulation setting with a good overall interobserver reliability. The majority of component steps showed a good interobserver agreement, as well as the general performance skills. By comparing the experienced with the inexperienced participants, significant differences in performance were found, establishing construct validity of this assessment tool for use on a simulation model. 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 toward a score goal of 30 (out of a possible total of 36). By adding the intermediate group, which was 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

| Table 3 Consistency (intraobserver and interobserver variability) between the scores given by expert A and expert B |
|-------------------------------------------------|-------------------------------------------------|------------------|------------------|
| Total score general skills<sup>a</sup> | Expert A (n = 82) | Expert B (n = 82) | ICC | p-Value |
| 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<sup>b</sup> | 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<sup>c</sup> | 26 (3.7; 0.4) | 26 (3.4; 0.4) | 0.669 | < 0.001 |

Abbreviation: ICC, intraclass correlate coefficient.
Note: Values depicted as mean with standard deviation and standard error of the mean, calculated with the ICC, on a two-tailed significance level of p < 0.05.
<sup>a</sup>General skills: sum of the tissue handling and instrument handling.
<sup>b</sup>Specific skills: 5 steps of the posterior sagittal anorectoplasty (PSARP).
<sup>c</sup>Overall: sum of general skills and specific skills.

| Table 4 Scores of inexperienced, intermediate and experienced group to determine construct validity |
|-------------------------------------------------|-------------------------------------------------|------------------|------------------|
| Total score general skills<sup>a</sup> | Inexperienced (n = 21) | Intermediate (n = 44) | Experienced (n = 16) | p-Value |
| 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<sup>b</sup> | 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<sup>c</sup> | 23.5 (3.1) | 25.4 (1.8) | 29.7 (2.7) | < 0.001 |

Note: Values depicted as mean with standard deviation, differences calculated with a one-way analysis of variance, on a two-tailed significance level of p < 0.05.
<sup>a</sup>General skills: sum of the tissue handling and instrument handling.
<sup>b</sup>Specific skills: 5 steps of the posterior sagittal anorectoplasty (PSARP).
<sup>c</sup>Overall: sum of general skills and specific skills.
surgery trainees throughout their training on simulation models.

A poor interobserver agreement was found for the steps “opening in midline” and “dissection of the fistula”; however, this could be related to the difficulty in simulating these steps on this low-budget PediaTrickBoxx ARM simulation model. Currently this is the only validated, inanimate ARM model available. Possibly, a high fidelity model would be able to simulate the difficulty of these steps even better, leading to a higher interobserver agreement. However, the increased cost might limit its usability in clinical practice. 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 among the trainees which may result in an improvement of the intra-observer correlation.

Overall, this CAT-PSARP had slightly lower interobserver reliability scores compared with the other previously developed CAT’s. 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 10 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 alone. Therefore, general skills and task-specific skills were weighted almost equally in the total score of overall performance. It has previously been suggested that procedure-specific assessments are more useful for trainees than a general assessment. Ahmed et al stated that the combination of global skills (e.g., tissue and instrument

|                          | Inexperienced versus intermediate | Intermediate versus experienced | Inexperienced versus experienced |
|--------------------------|----------------------------------|---------------------------------|---------------------------------|
| Total score general skills\(^a\) | 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\(^b\) | 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\(^c\) | 0.003                            | < 0.001                         | < 0.001                         |

Note: Inexperienced, \(n=21\); intermediate, \(n=44\); experienced, \(n=16\).

\(^a\)General skills: sum of the tissue handling and instrument handling.

\(^b\)Specific skills: 5 steps of the posterior sagittal anorectoplasty (PSARP).

\(^c\)Overall: sum of general skills and specific skills.

Fig. 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.
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 a general scoring system, only evaluating general skills.\textsuperscript{7}

**Limitations**
The assessment was done while participants practiced the PSARP procedure on a simulation model. The next step is to validate this assessment tool in the 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.\textsuperscript{25}

**Conclusion**
The CAT-PSARP showed a good overall interobserver agreement, as well as a good construct validity, for use on a simulation model 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 the most relevant in the assessment of the repair of an ARM by posterior sagittal approach.

**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.\textsuperscript{26} The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

**Availability of Data and Materials**
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Authors’ Contributions**
Study concept and design: M.J., I.B., and S.M.B.I.B.
Data acquisition: M.J., G.M.J.B., J.S., M.A.L., K.D., C.A.R., W.K., I.B., and S.M.B.I.B.
Analysis and data interpretation: M.J.
Drafting manuscript: M.J.
Critical revision: G.M.J.B., J.S., M.A.L., K.D., C.A.R., W.K., I.B., and S.M.B.I.B.
All authors read and approved the final manuscript.

**Funding**
The authors 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.

**Conflict of Interest**
The authors declare that they have no conflict of interest to declare.

**References**
1 Baldwin PJ, Paisley AM, Brown SP. Consultant surgeons’ opinion of the skills required of basic surgical trainees. Br J Surg 1999;86(08):1078–1082
2 Brennan MF, Debas HT. Surgical education in the United States: portents for change. Ann Surg 2004;240(04):565–572
3 Collins JP, Civil ID, Sugrue M, Balogh Z, Chehade MJ. Surgical education and training in Australia and New Zealand. World J Surg 2008;32(10):2138–2144
4 Frank JR, Langer B. Collaboration, communication, management, and advocacy: teaching surgeons new skills through the CanMEDS Project. World J Surg 2003;27(08):972–978, discussion 978
5 Smith AJ, Aggarwal R, Warren OJ, Paraskeva P. Surgical training and certification in the United kingdom. World J Surg 2009;33(02):174–179
6 Epstein RM. Assessment in medical education. N Engl J Med 2007;356(04):387–396
7 Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. Am J Surg 2011;202(04):469–480.e6
8 Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg 2001;193(05):479–485
9 Aggarwal R, Grantcharov TP, Eriksen JR, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg 2006;244(02):310–314
10 Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012;99(10):1322–1330
11 Eriksen JR, Grantcharov T. Objective assessment of laparoscopic skills using a virtual reality simulator. Surg Endosc 2005;19(09):1216–1219
12 Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. Br J Surg 2004;91(12):1549–1558
13 Ganni S, Chmarra MK, Goossens RHM, Jakimowicz JJ. Self-assessment in laparoscopic surgical skills training: is it reliable? Surg Endosc 2005;19(09):1216–1219
14 Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD. Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. Clin Orthop Relat Res 2016;474(04):874–881
15 Miskovic D, Ni M, Wyles SM, et al; National Training Programme in Laparoscopic Colorectal Surgery in England. 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(03):476–482
16 Ganni S, Botden SMBI, Schaap DP, Verhoeven BH, Goossens RHM, Jakimowicz JJ. “Reflection-before-practice” improves self-assessment and end-performance in laparoscopic surgical skills training. J Surg Educ 2018;75(02):527–533
17 IJgosse WM, Leijte E, Ganni S, et al. Competency assessment tool for laparoscopic suturing: development and reliability evaluation. Surg Endosc 2020;34(07):2947–2953
18 Levitt MA, Peña A. Anorectal malformations. Orphanet J Rare Dis 2007;2:33
19 Peña A, Devries PA. Posterior sagittal anorectoplasty: important technical considerations and new applications. J Pediatr Surg 1982;17(06):796–811
20 van Ling JA, Bökkerink GMJ, de Blaauw I, Botden SMBI. Development of a posterior sagittal anorectal surgical teaching model. BMC Pediatr 2021;21(01):57
21 Joosten M, Bökkerink GMJ, Levitt MA, et al. The use of an inanimate simulation model for the correction of an anorectal malformation in the training of colorectal pediatric surgery. Eur J Pediatr Surg 2021 (e-pub ahead of print). Doi: 10.1055/s-0041-1723035
22 Groenier M, Brummer L, Bunting BP, Gallagher AG. Reliability of observational assessment methods for outcome-based assessment of surgical skill: systematic review and meta-analyses. J Surg Educ 2020;77(01):189–201
23 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(02):155–163
24 Popović ZB, Thomas JD. Assessing observer variability: a user’s guide. Cardiovasc Diagn Ther 2017;7(03):317–324
25 Mahtani K, Spencer EA, Brassey J, Heneghan C. Catalogue of bias: observer bias. BMJ Evid Based Med 2018;23(01):23–24
26 WMO. Wet medisch-wetenschappelijk onderzoek met mensen. Accessed on April 16, 2021 at https://wetten.overheid.nl/BWB R0009408/2020-01-01