

University of Groningen

Procedure-based assessment for laparoscopic cholecystectomy can replace global rating scales

van Zwieten, Tom H.; Okkema, Sietske; Kramp, Kelvin H.; de Jong, Kim; Van Det, Marc J.; Pierie, Jean-Pierre E. N.

Published in:
Minimally invasive therapy & allied technologies

DOI:
[10.1080/13645706.2021.1995000](https://doi.org/10.1080/13645706.2021.1995000)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
van Zwieten, T. H., Okkema, S., Kramp, K. H., de Jong, K., Van Det, M. J., & Pierie, J-P. E. N. (2022). Procedure-based assessment for laparoscopic cholecystectomy can replace global rating scales. *Minimally invasive therapy & allied technologies*, 31(6), 865-871. <https://doi.org/10.1080/13645706.2021.1995000>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Procedure-based assessment for laparoscopic cholecystectomy can replace global rating scales

Tom H. van Zwieten, Sietske Okkema, Kelvin H. Kramp, Kim de Jong, Marc J. Van Det & Jean-Pierre E. N. Pierie

To cite this article: Tom H. van Zwieten, Sietske Okkema, Kelvin H. Kramp, Kim de Jong, Marc J. Van Det & Jean-Pierre E. N. Pierie (2021): Procedure-based assessment for laparoscopic cholecystectomy can replace global rating scales, *Minimally Invasive Therapy & Allied Technologies*, DOI: [10.1080/13645706.2021.1995000](https://doi.org/10.1080/13645706.2021.1995000)

To link to this article: <https://doi.org/10.1080/13645706.2021.1995000>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 26 Oct 2021.



Submit your article to this journal [↗](#)



Article views: 95



View related articles [↗](#)



View Crossmark data [↗](#)

Procedure-based assessment for laparoscopic cholecystectomy can replace global rating scales

Tom H. van Zwieten^{a,b}, Sietske Okkema^a, Kelvin H. Kramp^a, Kim de Jong^a, Marc J. Van Det^c and Jean-Pierre E. N. Pierie^{a,b}

^aDepartment of Surgery, Medical Centre Leeuwarden, Leeuwarden, The Netherlands; ^bPostgraduate School of Medicine, University Medical Centre Groningen, Groningen, The Netherlands; ^cDepartment of Surgery, Hospital Group Twente, Almelo/ Hengelo, The Netherlands

ABSTRACT

Introduction: Global rating scales (GRSs) such as the Objective Structured Assessment of Technical Skills (OSATS) and Global Operative Assessment of Laparoscopic Surgery (GOALS) are assessment methods for surgical procedures. The aim of this study was to establish construct validity of Procedure-Based Assessment (PBA) and to compare PBA with GRSs for laparoscopic cholecystectomy.

Material and methods: OSATS and GOALS GRSs were compared with PBA in their ability to discriminate between levels of performance between trainees who can perform the procedure independently and those who cannot. Three groups were formed based on the number of procedures performed by the trainee: novice (1–10), intermediate (11–20) and experienced (>20). Differences between groups were assessed using the Kruskal–Wallis and Mann–Whitney U tests.

Results: Increasing experience correlated significantly with higher GRSs and PBA scores (all $p < .001$). Scores of novice and intermediate groups overlapped substantially on the OSATS ($p = .1$) and GOALS ($p = .1$), while the PBA discriminated between these groups ($p = .03$). The median score in the experienced group was higher with less dispersion for PBA (97.2[85.3–100]) compared to OSATS (82.1[60.7–100]) and GOALS (80[60–100]).

Conclusion: For assessing skill level or the capability of performing a laparoscopic cholecystectomy independently, PBA has a higher discriminative ability compared to the GRSs.

ARTICLE HISTORY

Received 26 April 2021
Accepted 8 October 2021



KEYWORDS

Laparoscopic cholecystectomy; procedure-based assessment (PBA); key steps; construct validity; summative assessment

Introduction

Laparoscopic cholecystectomy is one of the most frequently performed surgical procedures worldwide and is commonly taught in the first three years of surgical training. Traditionally, residents are trained according to the master-apprentice model, in which a consultant surgeon (the master) provides side-by-side training and feedback for the resident (the apprentice) until the consultant surgeon declares that the resident is capable of performing a certain operation independently. To increase the reliability and reproducibility of the qualitative assessment of surgical trainees, several assessment tools have been developed. Martin et al.

[1] developed the Objective Structured Assessment of Technical Skills (OSATS) global rating scale in which quality of technical surgical skills – Respect for Tissue, Time and Motion, Instrument Handling, Knowledge of Instruments, Use of Assistants, Flow of Operation and Knowledge of Specific Procedure – are rated on a 5-point Likert scale. The OSATS has been validated in multiple studies and has become the gold standard for structured feedback in surgical training in the Netherlands [2–4]. Specifically for laparoscopic surgery, which requires some unique technical skills, Vassilliou et al. [5] developed the Global Operative Assessment of Laparoscopic Surgery (GOALS) to rate general laparoscopic skills – Depth Perception,

CONTACT Tom H. van Zwieten  thvanzwieten@gmail.com  Department of Surgery, Henri Dunantweg 2, Leeuwarden 8934AD, The Netherlands
On behalf of the expert panel representing the North-East Surgical School of the Netherlands: H.J.van Loon, Hospital Group Twente, W.A. ten Cate, Hospital Group Twente, K.P. Wevers, Medical Centre Leeuwarden, E.R. Manusama, Medical Centre Leeuwarden, M.A. Kaijser, Medical Centre Leeuwarden, M. Emous, Medical Centre Leeuwarden, F.A. Dijkstra, University Medical Centre Groningen, R.J. de Wit, Medisch Spectrum Twente, W.R. Hogeboom, Medisch Spectrum Twente, E. van de Krol, Medisch Spectrum Twente, A.E. Dassen, Medisch Spectrum Twente, F.W.H. Kloppenberg, Treant Care Group Emmen, B.P.J.A. Keller, Martini Hospital Groningen

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Bimanual Dexterity, Efficiency, Tissue Handling and Autonomy – on a 5-point Likert scale.

GRSs such as the OSATS and GOALS use subjective measures on general technical skills and can therefore be used for feedback and discussion (i.e., formative assessment). However, GRSs are not suitable for the examination or credentialing (i.e., summative assessment) for specific procedures, because they lack procedure specificity and related cut-off scores [4,6–8].

There is an increasing need for transparent and objective quality assessment and registration of trainee surgeons and the work they provide for their patients. The individual judgement of the ‘master’ is no longer sufficient. Moreover, due to working hour regulations and rotations through teaching hospitals with growing surgical departments, residents are often trained by multiple supervising surgeons who may be unaware of their portfolio and progression on a learning curve. To optimize surgical training, structured feedback per procedure is therefore needed. The PBA (procedure-based assessment) tool has been proposed as an alternative to the GRSs in assessing surgical performance, which enables clinicians to provide procedural-specific feedback and could facilitate examination in the performance of a procedure [8].

A PBA assessment form can be composed in several ways. A variation of the OSATS extended with the assessment of technical steps of a specific procedure has been described in previous studies. A PBA published in 2013 by Glarner et al. formed the basis for the PBA used in this research [9–11]. The steps of laparoscopic colon resections were connected to a scale of independence [12]. The scale of independence uses the frequency of verbal guidance and takeovers to assess the quality of surgical skills. In 2014 the Delphi method was applied to a group of laparoscopic surgeons to reach a consensus on the key steps in laparoscopic cholecystectomy [13]. In 2016 our research group published a PBA that consisted of a modified version of the independence scale of Glarner which was attached to the key steps obtained through the Delphi method [12,14]. The PBA was evaluated together with the OSATS and GOALS on validity, reliability and support for implementation by letting seven surgeons and three senior surgical trainees (4–6 post-graduate year (PGY)) assess three blinded videos of residents performing a laparoscopic cholecystectomy. Higher inter-rater reliability (different assessors agreed with one another more closely) and validity was found for PBA than for GRSs, suggesting that assessment of a series of procedural key-steps, on

which consensus has been achieved, compels to look at specific elements of operative competence and thereby gives less room for subjectivity than GRSs [14,15]. There also was strong support for implementation. However, for further implementation, the PBA still needs to be validated in daily clinical practice.

The first aim of the present study was to establish construct validity for the PBA on laparoscopic cholecystectomy and to compare the discriminative ability of GRSs and PBA during the immediate post-procedural assessment of surgical trainees. The second aim was to evaluate whether cut-off values could be derived from PBA for the summative assessment of being capable of performing a laparoscopic cholecystectomy independently.

Material and methods

Study population

All surgical trainees and supervising surgeons from eight teaching hospitals in the North-East Surgical region of the Netherlands were invited to participate in this prospective registration study.

Potential candidates for assessment included all residents who were in their learning curve for laparoscopic cholecystectomy, or who had completed their learning curve and were qualified to perform the procedure independently.

Data from elective laparoscopic cholecystectomies were collected between November 2016 and July 2018. All participating hospitals performed at least 200 elective laparoscopic cholecystectomies yearly. Supervising surgeons and residents all received the same instruction by an oral presentation during an initiation meeting in each participating hospital.

A standardized curriculum is used for laparoscopic cholecystectomy training which is similar for all residents in the North-East Surgical region and is centrally coordinated in this region. Residents begin this training with a mandatory course in basic laparoscopic skills. All supervisors were qualified surgeons, meaning that they had performed at least 200 laparoscopic cholecystectomies.

Patients were asked for their informed consent for the cholecystectomy to be performed in a training setting by a resident and a supervising surgeon. Because this study involved procedures performed in the context of regular care and no patient data were used (only data provided by physicians), approval from the medical ethical committee was not required.

The number of procedures performed based on trainee portfolios was used to form groups. Based on

experience with the learning curve of laparoscopic cholecystectomy, we estimated the number of procedures needed to perform a laparoscopic cholecystectomy independently: novice (1–10 laparoscopic cholecystectomies performed), intermediate (11–20 laparoscopic cholecystectomies performed) and experienced (>20 laparoscopic cholecystectomies performed).

The Post Graduate Year (PGY) was not used as a criterion because it is not a reliable indicator of skills and experience for laparoscopic cholecystectomy in the Netherlands. This is because residents in the Netherlands rotate between university medical centres and non-university teaching hospitals, which means that their rotation programmes can differ.

Measurements

The OSATS, GOALS and PBA were subsequently filled in by the supervising surgeons of the residents that performed the laparoscopic cholecystectomy. The number of laparoscopic cholecystectomies the residents had performed was also filled in, based on their portfolio. The procedures to be assessed were chosen by the resident scheduled for the procedure. Prior to each procedure, the trainee and supervising surgeon agreed to participate in the assessment. Only elective procedures for uncomplicated cholelithiasis were used for this. Laparoscopic cholecystectomy for acute cholecystitis was excluded. If a procedure turned out perioperatively to be too difficult for the trainee and the supervisor took over, it was excluded from the analysis. For each included procedure, all three assessment methods (PBA, OSATS and GOALS) were completed in random order immediately after each procedure online. The online assessment forms were filled in anonymously and the researchers received output in an Excel file. For participating trainees, it was optional to receive filled-in forms for their administration.

Statistical analysis

To evaluate whether the OSATS, GOALS and PBA could discriminate between levels of experience and thereby levels of technical skills in performing laparoscopic cholecystectomy, the total scores were first converted into standardized percentage scores of the maximum achievable score. This was calculated as follows: $\text{percentage score} = \frac{[(\text{total score} - \text{minimum score}) / (\text{maximum score} - \text{minimum score})] \times 100\%}{}$. The maximum score on the PBA is 72 points when

all values were filled in. The PBA contains values which could be filled in as ‘not applicable’ when certain steps were not needed, such as adhesiolysis. The maximum possible score was corrected for steps which were filled in as ‘not applicable’. The lowest possible score on the PBA is 0. For the OSATS the maximum score is 35 and the lowest possible score is 7. For the GOALS the maximum score is 25 and the lowest possible score is 5.

Boxplots from the OSATS, GOALS and PBA were created to assess whether the scores could discriminate between these three levels of experience. The Kruskal-Wallis test was used to assess differences in scores between the three groups. In the case of a significant Kruskal-Wallis test, the Mann-Whitney test was performed to assess the difference between the novice and intermediate group, the intermediate and experienced group and the novice and experienced group. To assess the dispersion of the scores within the groups, the quartile coefficient of variation was calculated based on the quartile scores $((Q3 - Q1) / (Q3 + Q1))$. Construct validity was reached when PBA adequately measured progression and could differentiate between groups based on experience level. To assess whether the scores discriminate between trainees that received an assessment ‘capable to perform the procedure independently’ and those that do not, boxplots were created and a Mann-Whitney test was used. Additionally, the area under the curve was used to estimate the ability of OSATS, GOALS and PBA to discriminate for being independent on the laparoscopic cholecystectomy. A cut-off score was estimated for PBA by measuring sensitivity and specificity. No false positives were accepted (incorrectly assessed as an independent) and therefore a 100% specificity was used. SPSS version 25 (IBM Corp. Armonk, NY, USA) was used for all statistical analyses. p -value < .05 was considered statistically significant.

Results

In total, 40 cholecystectomies were registered for assessment of which five were excluded because of missing data. For analysis 35 cholecystectomies (beginner 14, intermediate ten, experienced eleven) were included, completed by 13 different participants and 16 different supervisors. Six participants participated more than once with different procedures and varying supervisors. Boxplots of the scores from the three groups are depicted in [Figure 1](#). All three assessment tools showed increasing scores with increasing

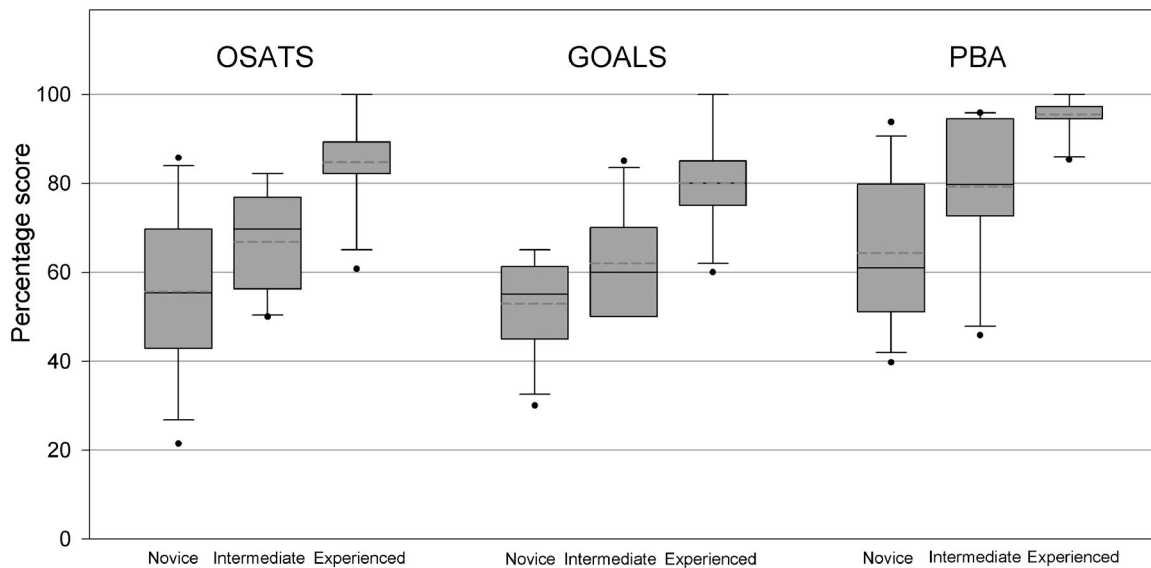


Figure 1. Scores from three groups showing substantial overlap between novice and intermediate group in OSATS and GOALS. PBA discriminates more effectively between all groups.

Table 1. Median (range) and quartile coefficient of dispersion (QCD) of the percentage score of assessment methods in three groups, with *p*-values for the difference between the novice versus intermediate and intermediate versus experienced groups.

		Novice (1–10)	Intermediate trainee (11–20)	Experienced trainee (>20)	<i>p</i> -value for novice vs. intermediate	<i>p</i> -value for intermediate vs. experienced
OSATS	Median (range)	55.4 (21.4–85.7)	69.6 (50.0–82.1)	82.1 (60.7–100)	<i>p</i> = .13	<i>p</i> < .001
	QCD	0.24	0.15	0.22		
GOALS	Median (range)	55.0 (30–65)	60.0 (55–85)	80.0 (60.0–100)	<i>p</i> = .10	<i>p</i> < .001
	QCD	0.15	0.17	0.13		
PBA	Median (range)	60.9 (39.7–93.8)	79.7 (45.8–95.8)	97.2(85.3–100)	<i>p</i> = .04	<i>p</i> < .001
	QCD	0.04	0.06	0.01		

The quartile coefficient of dispersion is a measure of dispersion and calculated as $(Q3 - Q1) / (Q3 + Q1)$.

experience levels (all $p < .001$, see Figure 1 and Table 1). Overall, the PBA scores were higher compared to the OSATS and GOALS. There was substantial overlap in scores between the novice and intermediate group for the OSATS ($p = .13$) and GOALS ($p = .10$), while the PBA was the only assessment that discriminated between these groups ($p = .04$) (Figure 1).

All three assessment tools were able to discriminate between the intermediate and experienced groups (Figure 1, all p -values < .001). The median score in the experienced group was closer to 100 with less dispersion for the PBA (median 97.2, range 85.3–100) compared to the OSATS (median 82.1, range 60.7–100) and GOALS (median 80.0, range 60.0–100) (Table 1).

On ten assessments, the supervisor stated that the trainee was able to perform the procedure without supervision. One assessment with the statement of capability to perform a laparoscopic cholecystectomy independently was in the intermediate group, nine assessments were in the experienced group. The

maximum number of laparoscopic cholecystectomies performed before being assessed as capable to perform the procedure independently was 35. On 25 assessments the supervisor stated that the trainee could not yet perform the procedure independently. The median OSATS score in the group that was able to perform the procedure without supervision was 83.9 (range 60.7–100). In the group that was unable to perform the procedure independently, the median score was 60.7 (range 21.4–85.7). The median total GOALS score was 75.0 (range 60.0–100) for the independent group and 55.0 (range 30.0–85.0) for the group that could not perform the procedure independently. The median total PBA score in the independent group was 97.2 (range 94.1–100) and 75.0 (range 39.7–95.8) in the group who could not perform the procedure independently (Table 2). A box-plot is shown in Figure 2.

The area under the curve (AUC) (95%CI) for OSATS was 0.898 (range 0.786–1.000), for GOALS this was 0.880 (range 0.765–0.995) and for PBA this

was 0.985 (range 0.956–1.000). When not accepting any false-positive measurements (incorrectly assessed as independent based on the score), a cut-off score for the PBA was suggested at 96.35. With this cut-off score, sensitivity was 70% and specificity 100%.

Discussion

Procedure-based assessment (PBA) is a method that rates trainees on their ability to perform key steps in a specific procedure by using an independence-based assessment model. In contrast, the OSATS and GOALS global rating scales (GRS) use subjective measures of general technical skills. Although these skills are essential for performing any surgical procedure, they do not reflect the independence, correctness and safety of the trainee's performance on the key steps that are needed to perform specific procedures. Therefore, we believe these GRSs are less suitable instruments for summative assessment.

In this multicentre study, we estimated the discriminative ability of GRSs and PBA by comparing

the scores of these assessment methods, filled in immediately after the trainee performed a laparoscopic cholecystectomy. The progression of the trainees in different groups was estimated and the assessment methods were compared. All three methods showed a significant difference between the intermediate and experienced groups. In particular, the statistically significant difference measured by the PBA between the novice and intermediate groups indicates that PBA is a more sensitive method and has more discriminative capability in the lower spectrum of scores than OSATS and GOALS. The difference in skills correlated directly with the level of experience. This supports the hypothesis that the PBA adequately measures progression, even to a larger extent than the OSATS and GOALS do, thus supporting the construct validity of the PBA. These results also support and extend findings in earlier studies that validated the OSATS and GOALS [2–4].

Overall, the scores on PBA were higher compared to OSATS and GOALS. This could be explained by a high score on assumed 'easy steps' at the beginning of

Table 2. Median (range) and quartile coefficient of dispersion (QCD) of the percentage score of all three assessment methods in trainees who could and could not independently perform a laparoscopic cholecystectomy.

		OSATS	GOALS	PBA
Not independent		60.7 (21.4–85.7)	55.0 (30.0–85.0)	75.0 (39.7–95.8)
	QCD	0.20	0.13	0.21
Independent		83.9 (60.7–100)	75.0 (60.0–100)	97.2 (94.1–100)
	QCD	0.06	0.12	0.01

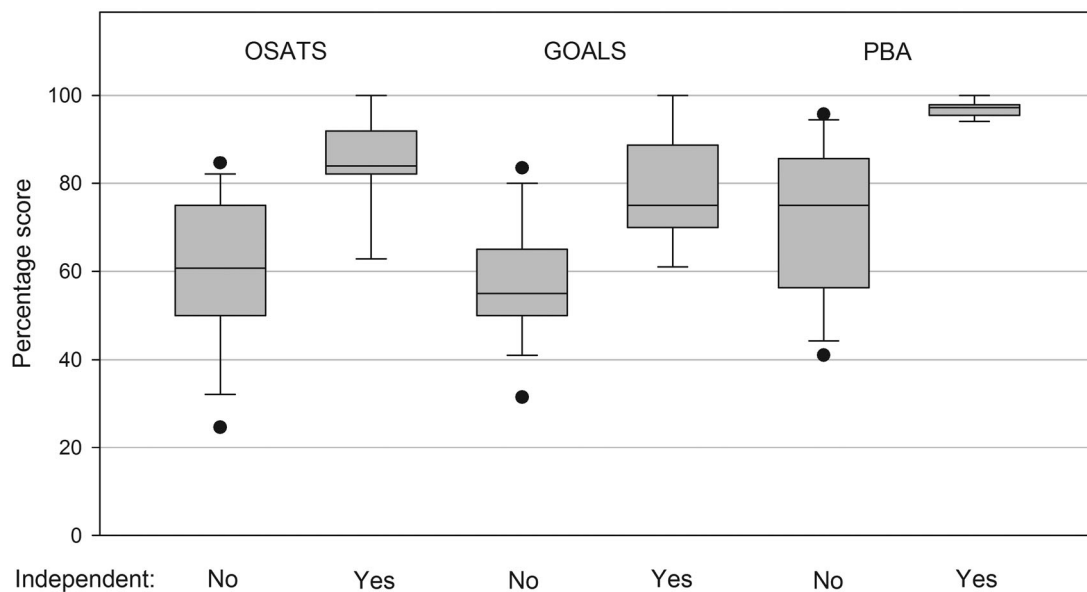


Figure 2. Scores from the three assessment methods for a trainee who is considered capable of performing a laparoscopic cholecystectomy independently. For the PBA, this score approached the maximum score with less dispersion compared to OSATS and GOALS.

the learning curve, which are scored in the PBA such as ‘positioning and port insertion’ and ‘ending the operation’. The wide dispersion in the novice and intermediate groups which is seen in the scores from the PBA could also be explained by this. Trainees who have broad experience with assisting other laparoscopic procedures or performing other laparoscopic procedures such as laparoscopic appendectomy will probably score maximum on these steps at the beginning of their learning curve. Also, it is conceivable that a maximum score on OSATS or GOALS is not necessary, because e.g., extensive experience is needed for optimal efficiency in OSATS.

In the Netherlands, residents rotate between teaching hospitals and are assessed by different supervisors. One of the main goals of developing and introducing the PBA for surgical procedures is to increase the objectivity of assessments made by different supervisors at different teaching hospitals. A total of 16 supervisors participated in this study, which minimised the possible effect of one supervisor being stricter than another. In our *post-hoc* analysis, the scores in different groups were indeed comparable without outliers.

Multiple studies have shown that GRSs can be used for formative assessment, but that they are not reliable for distinguishing between different levels of performance. They also lack a cut-off value for summative assessment. Therefore, GRSs cannot be used for a summative assessment [4,6,8].

For trainees who are considered capable of independently performing a laparoscopic cholecystectomy, the PBA score approaches the maximum score with little dispersion. In contrast to the OSATS and the GOALS, the PBA showed little overlap between trainees who were not yet capable of performing the procedure independently and those who were capable. To support the supervising surgeon, cut-off scores could be estimated between the trainees who could and could not yet independently perform the laparoscopic cholecystectomy. Although numbers are small, these results suggest that a minimum PBA score of 96 can be used as a cut-off score to predict whether a trainee is capable of independently performing a laparoscopic cholecystectomy (i.e., summative assessment). We have chosen to maximize the specificity because an incorrect (false-positive) estimate of being capable of performing the operation independently is undesirable. This would imply that a resident was allowed to perform a procedure without mastering all key steps. Determining a cut-off score should however be

subject to further research, including also more complex laparoscopic procedures.

Limitations

This study has several limitations that should be noted. First, numbers were small, and validation in a different sample with higher numbers of cases is clearly needed to confirm findings within this study. Nevertheless, in our study, the PBA showed a high discriminative capability based on a large area under the curve (AUC) and a small confidence interval with a high lower limit.

Another limitation of the current study is that many different trainees participated with one or more assessments, with a maximum of ten. Due to this limitation on the number of observations in the analysis, all assessments were considered to be independent. This assumption was supported by two sensitivity analyses that included only the first assessment (least experienced) or the last assessment (most experienced) of the trainee. However, the number of independent observations in the sensitivity analyses was small ($n = 13$).

Also, factors that could be of influence on the score, such as experience with assisting on laparoscopic procedures or experience with laparoscopic appendectomies e.g., are not accounted for in this research. This, however, is a limitation for all assessment procedures.

Despite the limitations of our study, we believe the PBA is a useful tool for supervisors to assess whether a trainee can perform a laparoscopic cholecystectomy independently. However, reaching the cut-off score on one assessment does not necessarily mean that the trainee can operate independently from then on. In the Netherlands current consensus is that a trainee needs three OSATS assessments of ‘independent’ from at least two different supervisors. The same applies to the PBA: the cut-off score should be reached repeatedly with different supervisors. Consensus should be reached on this before the PBA is implemented in practice.

Conclusion

The results of this study support the idea that PBA may be an alternative or additional assessment method to the commonly used GRSs for laparoscopic cholecystectomy. In contrast to GRSs, the PBA could differentiate between all skill levels, especially between trainees in the novice and intermediate phases of

learning. The PBA also appeared to have a high discriminative capability for assessing whether a trainee is capable of independently performing a laparoscopic cholecystectomy. We, therefore, believe that the PBA is a viable candidate in the summative assessment of surgical training.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

- [1] Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84:273–278.
- [2] Niitsu H, Hirabayashi N, Yoshimitsu M, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today*. 2013;43(3):271–275.
- [3] Hopmans CJ, den Hoed PT, van der Laan L, et al. Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. *Surgery*. 2014; 156(5):1078–1088.
- [4] van Hove PD, Tuijthof GJ, Verdaasdonk EG, et al. Objective assessment of technical surgical skills. *Br J Surg*. 2010;97(7):972–987.
- [5] Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190(1): 107–113.
- [6] Hiemstra E, Kolkman W, Wolterbeek R, et al. Value of an objective assessment tool in the operating room. *Can J Surg*. 2011;54(2):116–122.
- [7] Hatala R, Cook DA, Brydges R, et al. Constructing a validity argument for the Objective Structured Assessment of Technical Skills (OSATS): a systematic review of validity evidence. *Adv Health Sci Educ Theory Pract*. 2015;20(5):1149–1175.
- [8] Beard JD, Marriott J, Purdie H, et al. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health technol assess*. 2011;15:i–xxi.
- [9] Nickel F, Hendrie JD, Stock C, et al. Direct observation versus endoscopic video recording-based rating with the objective structured assessment of technical skills for training of laparoscopic cholecystectomy. *Eur Surg Res*. 2016;57(1–2):1–9.
- [10] Sarker SK, Chang A, Vincent C, et al. Development of assessing generic and specific technical skills in laparoscopic surgery. *Am J Surg*. 2006;191(2): 238–244.
- [11] Willuth E, Hardon SF, Lang F, et al. Robotic-assisted cholecystectomy is superior to laparoscopic cholecystectomy in the initial training for surgical novices in an ex vivo porcine model: a randomized crossover study. *Surg Endosc*. 2021;[e-pub ahead of print Feb 26].
- [12] Glarner CE, McDonald RJ, Smith AB, et al. Utilizing a novel tool for the comprehensive assessment of resident operative performance. *J Surg Educ*. 2013; 70(6):813–820.
- [13] Bethlehem MS, Kramp KH, van Det MJ, et al. Development of a standardized training course for laparoscopic procedures using Delphi methodology. *J Surg Educ*. 2014;71(6):810–816.
- [14] Kramp KH, et al. Validity, reliability and support for implementation of independence-scaled procedural assessment in laparoscopic surgery. *Surg Endoscopy* 2016;30(6):2288–2300.
- [15] Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Med Educ*. 2012;46(1):28–37.