



Development of a rating scale for objective assessment of performance in laparoscopic appendectomy surgery

Pramudith Sirimanna ^{*}, Stephen Boyce[†], Prashanth Gunanayagam^{*}, Marc A. Gladman [‡] and Vasi Naganathan^{*§}

^{*}Faculty of Medicine and Health, Concord Clinical School, University of Sydney, Sydney, Australia

[†]Oxford University Hospitals NHS Trust, Oxford, UK

[‡]Adelaide Medical School, University of Adelaide, Adelaide, Australia and

[§]Centre for Education and Research on Ageing, Concord Hospital and University of Sydney, Sydney, Australia

Key words

Delphi technique, laparoscopic appendectomy, rating scales, surgical education, technical skills assessment.

Correspondence

Pramudith Sirimanna, Departments of Surgery, Concord Repatriation General Hospital, Hospital Road, Concord, NSW 2139, Australia.
Email: pramsirimanna@gmail.com

P. Sirimanna PhD, FRACS, MBBS, BSc;

S. Boyce MScB, FRCS; **P. Gunanayagam** MBBS, BSc; **M. A. Gladman** PhD, FRACS;

V. Naganathan PhD, FRACP.

The authors would like to acknowledge that while the Delphi technique is referenced throughout the article and the authors stand by that method as appropriate for this type of analysis, as well as the conclusions presented, reaching consensus after a single iteration means that the presented research itself does not qualify as a Delphi study and should have more correctly been referred to as a 'ranking by experts'. The authors apologise for this oversight in preparing the article and for any misunderstanding that may have been caused.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Accepted for publication 25 February 2022.

doi: 10.1111/ans.17601

Introduction

The landscape of surgical education is changing. In 2022, the General Surgery Education and Training program (GSET) will be

moving towards a competency-based paradigm. Within this new curriculum, the focus will shift from volume of procedures performed to demonstration of surgical competence at a number of Procedure Based Activities (PBAs) in order to allow progression.

Abstract

Backgrounds: Laparoscopic appendectomy (LA) is the most common index procedure for junior surgical trainees. Despite the shift towards competency-based training, there is no method of quantitatively assessing performance during LA. This study aimed to obtain expert consensus regarding the items required to create a LA Rating Scale (LARS).

Methods: A list of steps required for LA surgery, as well as descriptors of "poor", "average" and "excellent" performance for each of these steps were created for potential inclusion into an objective assessment tool for LA surgery. Using a Delphi method, 20 experts from multiple institutions rated on a Likert scale from 1 to 5 the suitability of these steps and descriptors of performance. Responses were obtained until consensus (Cronbach's $\alpha > 0.8$) was achieved.

Results: Fifteen experts participated in the study. Consensus was achieved for all items during the first iteration of the Delphi with a Cronbach's α of 0.96. The Cronbach's α for the steps was 0.87 and 0.92 for the descriptors of performance. Steps and descriptors of performance that $>80\%$ of experts rated as ≥ 4 were used to create the final LARS tool.

Conclusion: Multi-institutional expert consensus was obtained regarding the steps and, for the first time, descriptors of performance for LA, demonstrating their face and content validity, as well as generalisability. Subsequently, the LARS tool was created that can be used to quantitatively assess intra-operative performance. This instrument can be used to identify weaknesses in performance and facilitate deliberate practice, thus shifting training in LA to a competency-based approach.

As an example, in the recently published GSET regulations, appendectomy is deemed a core procedure and one that trainees must be able to complete independently with minimal supervision and guidance by the end of their third year of training.¹ Indeed, the laparoscopic appendectomy (LA) is the principal index procedure performed by surgical trainees. In many instances, it provides the arena for a trainee's first endeavour into the world of laparoscopic surgery as the primary operator. Furthermore, it is associated with a notable learning curve of between 20 and 30 cases, during which errors are most likely.^{2,3}

Whilst PBAs have an important place in a competency-based training system, the method of assessing performance and competence during these PBAs must be objective, standardized and give structured feedback on which areas trainees require further training. A number of observational evaluation tools have been developed to assess surgical performance in an objective and standardized manner providing a foundation for constructive feedback.⁴ This method of assessment focuses on the use of rating scales to generate a quantitative measure of procedural performance through direct observation. These tools exist in two broad categories; global rating scales that assess generic skills applicable to any operation, and procedure-specific rating scales that assess skills specific to a particular procedure.

Although global rating scales, such as the Objective Structured Assessment of Technical Skill (OSATS), have been widely demonstrated as valid and reliable in both open and laparoscopic surgery within the simulated and actual operating theatre environment, they provide little feedback on performance of the specific operative steps of the procedure.⁴⁻⁹ The value of assessment of competency in specific steps is that the trainee then has feedback on which steps they have achieved proficiency in, and which steps require more training. As such, procedure-specific rating scales have been developed and validated for a number of surgical procedures, including laparoscopic cholecystectomy, bariatric and colorectal surgery.¹⁰⁻¹³ Not only do the evaluations of performance generated by these assessment tools provide an objective assessment of competence, but they can also be used to grant privileges for independent practice to surgeons who have learnt new procedures.⁴ Most of these evaluation tools have been developed within institutions using faculty opinion. There are, however, benefits to utilizing a more systematic approach to developing an evaluation tool with input across a variety of institution.¹²

Thus far, an evidence-based systematically derived tool to evaluate performance during a LA has yet to be constructed. Such a tool would allow trainees to have their performance objectively assessed to provide meaningful constructive feedback so that their rate of skill acquisition can be monitored, and weaknesses practiced deliberately prior to entering independent practice. Thus, the aim of this study was to develop a procedure-specific evaluation tool for assessment of performance in LA surgery.

Methods

Study design

A Delphi method was used in order to obtain consensus amongst a group of surgeon experts with regards to the essential steps for a

laparoscopic appendectomy, as well as descriptors of what constitutes "poor", "average" and "excellent" performance at each of these steps. Developed by the RAND Corporation,¹⁴ the premise of a Delphi method is to allow the thoughts and opinions of individuals to be gathered and assessed repeatedly until consensus is reached within the group.¹⁵ The Delphi process has been shown to be feasible, low cost and reliable.¹⁵ Anonymised questionnaires are used in the Delphi process which means participants do not need to physically meet, and this also ensures that dominant participants do not influence the outcome of the group.

Selection of expert surgeons

Twenty surgeons in General Surgery were selected to participate in the Delphi process on the basis that they were known to have an interest in education as demonstrated by their involvement in professional organizations such as the Royal Australasian College of Surgeons, the Academy of Surgical Educators, the Society of Gastrointestinal and Endoscopic Surgeons, and other national surgical societies, as well as those with a strong academic and professional record in surgical education. Involvement of the participants into the expert panel was anonymised.

Delphi method

A list of potential steps and sub-steps required to perform a LA were derived using surgical textbooks and peer-reviewed literature, as well as gaining the opinion of faculty experts at Concord Repatriation General Hospital. Additionally, descriptors of "poor", "average" and "excellent" performance were formulated for each of sub-steps by two faculty experts. Potential errors during LA surgery were included within descriptions of "poor" performance. Each participant was then emailed a link to an online questionnaire using Survey Monkey™ (Palo Alto, CA) that asked them to rate each step and sub-step with respect to how strongly the participants felt they should be included in the final evaluation tool using a Likert scale from 1 (strongly disagree) to 5 (strongly agree). Additionally, participants were asked to rate the accuracy of the descriptors of "poor", "average" and "excellent" performance for each sub-step using the same Likert scale. Email reminders were sent after 2 weeks and 1 month after the initial questionnaire was sent to those who were yet to complete the questionnaire.

Determining consensus and creation of the evaluation tool

There are no recognized criteria for establishing consensus when performing a Delphi survey. However for our study, based on work by Graham et al. and Palter et al., we used Cronbach's α as a measure of the internal consistency of responses and used this as an indicator of the degree of consensus amongst the expert panel.^{12,15} Greater internal consistency amongst a group of measures occurs as Cronbach's α approaches 1.0, however, there are a variety of cut-off values used within the literature to deem when consensus has been reached. Palter et al. suggested an α value of greater than 0.8 is suitable for the purposes of determining consensus for an evaluation tool.¹² Thus, for this study this the cut-off was chosen to

determine if consensus had been reached amongst the expert panel for the survey overall, and then individually for the steps/sub-steps and the descriptors of performance. After consensus was reached, the steps, sub-steps and descriptors of performance that at least 80% of experts either agreed or strongly agreed with were incorporated into the final evaluation tool.¹²

Data analysis

Mean and standard deviations (SDs) for scores given to each step, sub-step and descriptors of performance were calculated. Cronbach's α was calculated to determine the internal consistency, and thus consensus, of the responses of the expert panel. All analyses were performed using SPSS (Statistical Package for Social Sciences version 20.0, Chicago, IL, USA). Ethical approval for this study was obtained from the Human Research Ethics Committee—Concord Hospital of the Sydney Local Health District.

Results

Fifteen of the twenty expert surgeons (75%) contacted participated in the Delphi survey. Fourteen of expert participants were from Australia from seven different hospitals, whilst one was from the United States. Following collection of responses from the first round of the survey, the Cronbach's α for the entire survey was 0.96. The Cronbach's α for the step and sub-steps was 0.87 and for the descriptors of performance was 0.92. At least 80% of participants either agreed or strongly agreed that all steps and sub-steps should be included into the final LA Rating Scale (LARS) except "Perform thorough washout using aspirator" (Table 1). Similarly, 80% or more of participants either agreed or strongly agreed that descriptors of performance were accurate representations of "poor", "average" and "excellent" technical skill of the sub-steps, except

for descriptions of "Inspect all four quadrants" and "Perform thorough washout using aspirator" (Table 2). For the "Inspect all four quadrants" sub-step, only 53.3% of participants agreed or strongly agreed that the descriptors for poor and average performance were accurate. Similarly, for the "Perform thorough washout using aspirator" sub-step, 73.3% of participants either agreed or strongly agreed that the descriptors of average and excellent performance were accurate. Using these results, as well as comments from the expert panel, the final LARS tool was created (Figure 1).

Discussion

The development of tools to assess competency in surgery, particularly for an index procedure like laparoscopic appendectomy, is essential to allow objective and useful feedback for surgical trainees, as well as help training programs move towards competency-based system. In this study, we designed an evaluation tool that can be used to assess technical competence in LA surgery. This was done using a Delphi method by obtaining consensus amongst a group of expert surgeons with regards to the components to be included in the tool.

All steps and sub-steps that at least 80% of participants either agreed or strongly agreed should be included were integrated into the final evaluation tool. As only 66.7% of participants either agreed or strongly agreed that the 'Perform thorough washout using aspirator' sub-step should be included in the tool, and 73.3% either agreed or strongly agreed that the descriptors of 'average' and 'excellent' performance for this step were accurate, this step was not included in the final tool. Participants stated that this step was a duplicate of the 'Suction/Lavage any free fluid/pus' sub-step. All descriptors of performance that 80% or more of participants either agreed or strongly agreed were accurate representations of 'poor', 'average' and 'excellent' technical skill at each of the included sub-

TABLE 1 Results for assessment by the 15 participants of the suitability of the steps and sub-steps to be included in the final LARS tool

| Step/sub-step | Mean Likert score (0–5) | Likert score standard deviation | % Of experts who agreed or strongly agree the step should be included in LARS |
|---|-------------------------|---------------------------------|---|
| Diagnostic laparoscopy | 4.4 | 0.51 | 100 |
| Inspect of all four quadrants | 4.1 | 0.59 | 86.7 |
| Suction any free fluid | 4.5 | 0.52 | 100 |
| Exposure and mobilization of the appendix | 4.6 | 0.51 | 100 |
| Place patient in Trendelenberg position with right side up | 4.0 | 0.66 | 80 |
| Retract or Sweep the small bowel | 3.8 | 0.41 | 80 |
| If not immediately obvious, able to identify appendix using key anatomical landmarks | 4.4 | 0.63 | 93.3 |
| Divide peritoneal or inflammatory adhesions | 4.3 | 0.46 | 100 |
| Dissection or Division of mesoappendix | 4.3 | 0.62 | 93.3 |
| Grasp and retract appendix or mesoappendix to orientate and expose area for dissection | 4.3 | 0.62 | 93.3 |
| Dissection of the mesoappendix in order to create a mesenteric window or identify the appendicular artery or skeletonize the appendix | 3.9 | 0.52 | 80 |
| Divide mesoappendix or appendicular artery if relevant | 4.3 | 0.46 | 100 |
| Division of appendix | 4.3 | 0.45 | 100 |
| Assessment of the condition and appropriate division of appendix at the base | 4.2 | 0.41 | 100 |
| Removal of appendix | 4.3 | 0.49 | 100 |
| Place the appendix into a bag | 4.2 | 0.56 | 93.3 |
| Deliver the appendix through the umbilical port | 4.1 | 0.64 | 86.6 |
| Inspection of operative bed | 4.4 | 0.51 | 100 |
| Inspect operative bed for bleeding | 4.4 | 0.51 | 100 |
| Perform thorough wash out using aspirator | 3.6 | 0.63 | 66.67 |

TABLE 2 Step, sub-steps and descriptors of poor, average and excellent performance used within the Delphi survey

| Operative step | Description of performance, mean Likert score, (SD), % of participants that agreed or strongly agreed that the item was an accurate description of performance for the corresponding task | | | | |
|---|--|--|-------------|---|--|
| | 1 (poor) | 2 | 3 (average) | 4 | 5 (excellent) |
| Diagnostic laparoscopy | | | | | |
| Inspect all four quadrants with view to make diagnosis of appendicitis and inspect for other differentials | <i>Not performed</i> 3.3, (1.10) 53.5% | <i>Moderately detailed inspection, not all 4 quadrants visualised. Prompting required to inspect other diagnoses</i> *(3.5, 0.99, 53.5%) | | | <i>Careful and thorough inspection of all 4 quadrants. Inspection for other diagnoses without prompting, including complete examination of small bowel and relevant pelvic structures</i> *(4.1, 0.91, 80%) |
| Suction/Lavage any free fluid/pus | <i>Omission of suction/lavage</i> *(4.0, 0.66, 80%) | <i>Partially complete suctioning/lavage; Performed after guidance</i> *(3.9, 0.59, 80%) | | | <i>Safe and thorough suctioning without guidance. Lavage until fluid return clear</i> *(4.2, 0.56, 93.3%) |
| Exposure and mobilisation of appendix | | | | | |
| Appropriately position patient to aid exposure | <i>Not performed</i> *(3.7, 1.16, 80%) | <i>Performed with guidance, excessive/insufficient tilt</i> *(3.7, 0.59, 80%) | | | <i>Performed to appropriate degree of tilt, without guidance</i> *(4.27, 0.46, 100%) |
| Retract/Sweep the small bowel/omentum to aid exposure | <i>Performed clumsily with poor technique, unsure how to retract small bowel. Repeated guidance required</i> *(3.9, 0.26, 93.3%) | <i>Performed adequately but not efficiently with some guidance and frequent repositioning of bowel required</i> *(3.9, 0.52, 80%) | | | <i>Performed skillfully and efficiently with gentle handling and effective displacement of bowel. No guidance required</i> *(4.1, 0.64, 86.7%) |
| Identifies and exposes appendix using anatomical landmarks | <i>Struggles to identify appendix not making use key anatomical landmarks. Repeated guidance required</i> *(4.3, 0.49, 100%) | <i>Is able to identify appendix but requires occasional guidance</i> *(3.8, 0.78, 86.7%) | | | <i>Easily identifies appendix using key anatomical landmarks without guidance</i> *(4.3, 0.46, 100%) |
| Divide peritoneal/inflammatory adhesions of appendix +/- caecum and/or proximal right colon as required to mobilize and locate the base of the appendix | <i>Unable to perform; Dangerous dissection/consistently lacks respect for tissues; Damage to surrounding structures. Repeated guidance required</i> *(4.3, 0.62, 93.3%) | <i>Safe but uneconomical movements. Requires some guidance</i> *(4.0, 0.66, 80%) | | | <i>Performed skillfully with safe/economical movement with full respect for tissues and without guidance</i> *(4.3, 0.62, 93.3%) |
| Dissection/Division of mesoappendix | | | | | |
| Retract appendix/mesoappendix to orientate and expose area for dissection | <i>Appendix/mesoappendix repeatedly dropped; Trauma to Appendix/mesoappendix; Poor retraction; Unable to orientate; Repeated guidance required</i> *(4.1, 0.52, 93.3%) | <i>Appendix/mesoappendix occasionally dropped; Minimal trauma to appendix/mesoappendix; Adequate retraction; Able to orientate with some guidance</i> *(4.2, 0.41, 100%) | | | <i>Appendix/mesoappendix skillfully retracted with appropriate tension and no trauma; Able to orientate without guidance</i> *(4.4, 0.51, 100%) |
| Dissection of the mesoappendix | <i>Unable to identify dissection planes; Dangerous dissection; Damage to surrounding structures/vessels; Repeated guidance required</i> *(4.3, 0.46, 100%) | <i>Adequate dissection with occasional guidance; Awareness of vessels and surrounding structures with occasional guidance</i> *(4.3, 0.46, 100%) | | | <i>Skillful and efficient dissection with awareness of vessels and surrounding structures; Without guidance</i> *(4.2, 0.56, 93.3%) |
| Divide mesoappendix/appendicular artery if relevant (i.e. clips appendicular artery) | <i>Multiple attempts made to divide mesoappendix/artery with traumatic movements; trauma to artery. Repeated guidance required</i> *(4.1, 0.35, 100%) | <i>Mesoappendix/artery divided in correct position with multiple movements with occasional guidance</i> *(4.1, 0.35, 100%) | | | <i>Mesoappendix/artery divided at correct position with efficient movement without guidance</i> *(4.2, 0.41, 100%) |
| Division of appendix | | | | | |
| Assessment of the condition, and appropriate ligation and division of appendix at the base | <i>No evaluation of condition of base. Inadequate ligation of the base. No appreciation of correct stump length. Damage of surrounding structures/contamination of luminal contents. Repeated guidance required</i> *(4.3, 0.46, 100%) | <i>Evaluation of condition of base with guidance. Adequate ligation of the base with some difficulty and guidance. Safe division of the appendix without contamination of luminal contents through repeated actions and some guidance</i> *(4.2, 0.41, 100%) | | | <i>Full evaluation of condition of base. Adequate ligation of the base with ease and no guidance. Leave adequate stump. Keen awareness of adjacent structures. No contamination of luminal contents. No guidance required</i> *(4.3, 0.46, 100%) |

Table 2 Continued

| Operative step | Description of performance, mean Likert score, (SD), % of participants that agreed or strongly agreed that the item was an accurate description of performance for the corresponding task | | | | |
|---|---|---|---|---|--|
| | 1 (poor) | 2 | 3 (average) | 4 | 5 (excellent) |
| Removal of appendix | | | | | |
| Place the appendix into a bag | <i>Clumsy repeated and ineffectual movements to place the appendix into the bag. Actual contamination. Repeated guidance required</i> *(4.1, 0.64, 86.7%) | | <i>Appendix placed in the bag using a number of movements. Potential contamination. Occasional guidance</i> *(3.9, 0.74, 80%) | | <i>Skillful economical placement of the appendix into the bag. No contamination. No guidance</i> *(4.1, 0.64, 86.7%) |
| Deliver the appendix through the umbilical port | <i>Inappropriate use of force, tear to bag or loss of appendix. Actual contamination. Repeated guidance required</i> *(4.4, 0.51, 100%) | | <i>Appendix retrieved with occasional guidance</i> *(3.9, 0.52, 80%) | | <i>Appendix retrieved with appropriate use of traction without guidance/guidance</i> *(4.3, 0.46, 100%) |
| Inspection of operative bed | | | | | |
| Inspect operative bed for bleeding | <i>Not performed</i> *(4.2, 0.56, 93.3%) | | <i>Performed with guidance</i> *(3.9, 0.52, 80%) | | <i>Careful checking of operative field without guidance</i> *(4.3, 0.46, 100%) |
| Perform thorough wash out using aspirator | <i>Random use of lavage. Tissues caught in aspirator. Repeated prompting required. Omission of suction/lavage</i> *(3.9, 0.52, 80%) | | <i>Incomplete suctioning; Performed after prompting. Safe use of suction occasional prompting</i> *(3.8, 0.56, 73.3%) | | <i>Complete thorough suctioning without prompting. Careful use of suction avoiding tissues. No prompting</i> *(3.8, 0.56, 73.3%) |

Note: This table displays the results for the assessment of accuracy of the descriptors of performance by the participants using the Delphi survey.
*Mean, (SD), % of participants that agreed or strongly agreed that the item was an accurate description of performance for the corresponding task).

| Operative Step | Performance | | | | |
|---|--|---|--|---|--|
| | 1 | 2 | 3 | 4 | 5 |
| Diagnostic laparoscopy | | | | | |
| Inspection of all four quadrants and inspect for other differentials (including small bowel run) if appendix macroscopically normal | <i>Not performed.</i> | | | | <i>Performed</i> |
| Suction/Lavage any free fluid / pus | <i>Omission of suction/lavage.</i> | | <i>Partially complete suctioning/lavage; Performed after guidance.</i> | | <i>Safe and thorough suctioning without guidance. Lavage until fluid return clear.</i> |
| Exposure and mobilisation of appendix | | | | | |
| Appropriately position patient to aid exposure | <i>Not performed.</i> | | <i>Performed with guidance, excessive/insufficient tilt</i> | | <i>Performed to appropriate degree of tilt, without guidance</i> |
| Retract / Sweep the small bowel / omentum to aid exposure | <i>Performed clumsily with poor technique, unsure how to retract small bowel. Repeated guidance required.</i> | | <i>Performed adequately but not efficiently with some guidance and frequent repositioning of bowel required</i> | | <i>Performed skillfully and efficiently with gentle handling and effective displacement of bowel. No guidance required.</i> |
| Identifies and exposes appendix using anatomical landmarks | <i>Struggles to identify appendix not making use key anatomical landmarks. Repeated guidance required.</i> | | <i>Is able to identify appendix but requires occasional guidance</i> | | <i>Easily identifies appendix using key anatomical landmarks without guidance</i> |
| Divide peritoneal/inflammatory adhesions of appendix +/- caecum and/or proximal right colon as required to mobilise and locate the base of the appendix | <i>Unable to perform; Dangerous dissection / consistently lacks respect for tissues; Damage to surrounding structures. Repeated guidance required.</i> | | <i>Safe but uneconomical movements. Requires some guidance.</i> | | <i>Performed skillfully with safe / economical movement with full respect for tissues and without guidance</i> |
| Dissection / Division of mesoappendix | | | | | |
| Retract appendix / mesoappendix to orientate and expose area for dissection | <i>Appendix / mesoappendix repeatedly dropped; Trauma to Appendix / mesoappendix; Poor retraction; Unable to orientate; Repeated guidance required.</i> | | <i>Appendix / mesoappendix occasionally dropped; Minimal trauma to appendix / mesoappendix; Adequate retraction; Able to orientate with some guidance</i> | | <i>Appendix / mesoappendix skillfully retracted with appropriate tension and no trauma; Able to orientate without guidance.</i> |
| Dissection of the mesoappendix | <i>Unable to identify dissection planes; Dangerous dissection; Damage to surrounding structures/vessels; Repeated guidance required.</i> | | <i>Adequate dissection with occasional guidance; Awareness of vessels and surrounding structures with occasional guidance</i> | | <i>Skillful and efficient dissection with awareness of vessels and surrounding structures; Without guidance.</i> |
| Divide mesoappendix / appendicular artery if relevant (i.e. clips appendicular artery) | <i>Multiple attempts made to divide mesoappendix/artery with traumatic movements; trauma to artery. Repeated guidance required.</i> | | <i>Mesoappendix/artery divided in correct position with multiple movements with occasional guidance.</i> | | <i>Mesoappendix/artery divided at correct position with efficient movement without guidance.</i> |
| Division of appendix | | | | | |
| Assessment of the condition, and appropriate ligation and division of appendix at the base | <i>No evaluation of condition of base. Inadequate ligation of the base. No appreciation of correct stump length. Damage of surrounding structures/contamination of luminal contents. Repeated guidance required.</i> | | <i>Evaluation of condition of base with guidance. Adequate ligation of the base with some difficulty and guidance. Safe division of the appendix without contamination of luminal contents through repeated actions and some guidance.</i> | | <i>Full evaluation of condition of base. Adequate ligation of the base with ease and no guidance. Leave adequate stump. Keen awareness of adjacent structures. No contamination of luminal contents. No guidance required.</i> |
| Removal of appendix | | | | | |
| Place the appendix into a bag | <i>Clumsy repeated and ineffectual movements to place the appendix into the bag. Actual contamination. Repeated guidance required.</i> | | <i>Appendix placed in the bag using a number of movements. Potential contamination. Occasional guidance.</i> | | <i>Skillful economical placement of the appendix into the bag. No contamination. No guidance.</i> |
| Deliver the appendix through the umbilical port | <i>Inappropriate use of force, tear to bag or loss of appendix. Actual contamination. Repeated guidance required.</i> | | <i>Appendix retrieved with occasional guidance</i> | | <i>Appendix retrieved with appropriate use of traction without guidance/guidance</i> |
| Inspection of operative bed | | | | | |
| Inspect operative bed for bleeding | <i>Not performed</i> | | <i>Performed with guidance</i> | | <i>Careful checking of operative field without guidance</i> |

FIGURE 1. Final LARS tool.

steps were incorporated into the final evaluation tool, except for 'Inspect all four quadrants' sub-step. Whilst 100% of participants thought the 'Inspect all four quadrants' sub-step was important to include in the rating scale, only 53.3% of participants agreed or strongly agreed that the descriptors for 'poor' and 'average' performance were accurate. When examining the comments, the participants stated that this sub-step should only be performed if the diagnosis of appendicitis is unclear and thus, not performing it does not constitute poor performance if it is not indicated. Furthermore, it was suggested by the group that if the diagnosis is in doubt, the descriptors of performance should be binary. As a result, the sub-step description was changed to 'Inspection of all four quadrants and inspect for other differentials (including small bowel run) if appendix macroscopically normal', and the descriptors of performance was changed to 'not performed' and 'performed'. Consensus regarding the inclusion of a 'Access and port insertion' step has previously been obtained using a similar methodology and therefore was not included within the steps of laparoscopic appendectomy assessed for inclusion into this tool.¹² However, in practice, the tool can be modified to add this previously validated step.

The used of an unbiased technique like a Delphi method, and the involvement of surgeons from multiple institutions adds strength to the face and content validity of the tool, as well as the generalisability. The expert surgeons were contacted via email, the survey was delivered online and the lack of interaction between participants negated the effect of individual group members dominating and influencing the process. Although a similar method has been utilized in the creation of technical skills assessment tools, this is the first-time descriptors of performance for a specific surgical procedure have been evaluated using the Delphi survey method.¹²

Despite the benefits of using a Delphi method to gain expert opinion and consensus there are some limitations to our study. It is possible that the investigators of this study had too much of an influence in the creation of the evaluation tool since they created the original list of steps and sub steps and wrote the original procedure descriptors.¹⁵ To abate this, we used a number of sources to create the list of procedure steps and descriptors of the steps and included broad descriptions of the steps and sub-steps to ensure applicability to any specific technique. A variety of descriptors were also used to indicate 'poor', 'average' and 'excellent' performance allowing a more accurate assessment of their applicability to be included in the evaluation tool. Specifically, descriptions of potential errors during LA surgery were included in order to identify 'poor' and unsafe performance. We gave an opportunity for the participants to provide comments on each item of the survey. A possible drawback of this, however, was that the survey was quite long which may have explained why some participants did not chose to participate. Fifteen of the 20 experts who agreed to participate completed the first round of the Delphi survey reflecting a similar response rate to other studies within the literature.^{16,17} The panel represents the shared expertise from multiple institutions from Australia but the fact that there was only one person from outside Australia took part is a potential source of bias. Whilst Cronbach's α can be effected by the number of responses (with a decreasing number likely to decrease its value), consensus was reached within

the first iteration of the Delphi survey suggesting that the items included were an accurate representation of the steps no matter what specific technique the surgeons used, and importantly, the descriptors of performance were deemed an accurate assessment of skill level at each of the step.

This study successfully utilized a Delphi method to obtain consensus amongst a group of expert surgeons across multiple institutions regarding the step and sub-steps that should be included in a tool to evaluate performance during laparoscopic appendectomy surgery. Importantly, for the first time in the development of a tool to evaluate a surgical procedure, agreement with regards to descriptors of performance during these steps were also obtained. This allowed the systemic construction of an objective assessment tool with robust content validity and widespread applicability. Indeed, trainees can use this tool to not only gain a greater comprehension of the steps and performance standards required to excel at LA surgery, but also use it to facilitate their training and shorten the learning curve, potentially for the benefit of patient safety, by allowing trainees to understand their strengths and weakness and partake in deliberate practice. Furthermore, this tool can be beneficial to supervisors. It can be used to track trainees' progress through the learning curve, identifying those who require further training. Additionally, it has the potential to be used for credentialing and revalidation purposes, for example during PBAs within the GSET program, by ensuring minimum standards are reached prior to embarking upon unsupervised practice in LA surgery and maintained after.

Acknowledgements

Pramudith Sirimanna was funded by a National Health and Medical Research Council (NHMRC) postgraduate scholarship during the conduct of the study (Scholarship number GNT1093784). Marc Gladman has received support from Karl Storz and Applied medical, during the conduct of the study and personal fees from Health Ed, Elsevier, and Abbvie LTD, outside the submitted work. Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

Author contributions

Pramudith Sirimanna: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; validation; visualization; writing – original draft; writing – review and editing. **Stephen Boyce:** Conceptualization; data curation; methodology. **Prashanth Gunanayagam:** Conceptualization; data curation; formal analysis; investigation; methodology. **Marc A. Gladman:** Funding acquisition; methodology; supervision. **Vasi Naganathan:** Formal analysis; supervision; writing – original draft; writing – review and editing.

References

1. Surgeons ABIGSRACo. Training regulations. General Surgery Education and Training Program. In: Surgeons ABIGSRACo, (ed.); 2021.

2. Jaffer U, Cameron AEP. Laparoscopic appendectomy: a junior Trainee's learning curve. *J Soc Laparoendosc Surg* 2008; **12**: 288–91.
3. Kim SY, Hong SG, Roh HR, Park SB, Kim YH, Chae GB. Learning curve for a laparoscopic appendectomy by a surgical trainee. *J Korean Soc Coloproctol* 2010; **26**: 324–8.
4. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Ann Surg* 2008; **247**: 372–9. <https://doi.org/10.1097/SLA.0b013e318160b371>.
5. Martin JA, Regehr G, Reznick R *et al*. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; **84**: 273–8.
6. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med* 1996; **71**: 1363–5.
7. Datta V, Bann S, Beard J, Mandalia M, Darzi A. Comparison of bench test evaluations of surgical skill with live operating performance assessments. *J Am Coll Surg* 2004; **199**: 603–6.
8. Anastakis DJ, Regehr G, Reznick RK *et al*. Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg* 1999; **177**: 167–70.
9. Hance J, Aggarwal R, Stanbridge R *et al*. Objective assessment of technical skills in cardiac surgery. *Eur J Cardiothorac Surg* 2005; **28**: 157–62. <https://doi.org/10.1016/j.ejcts.2005.03.012>.
10. Sarker SK, Chang A, Vincent C, Darzi SA. Development of assessing generic and specific technical skills in laparoscopic surgery. *Am J Surg* 2006; **191**: 238–44.
11. Sarker SK, Kumar I, Delaney C. Assessing operative performance in advanced laparoscopic colorectal surgery. *World J Surg* 2010; **34**: 1594–603. <https://doi.org/10.1007/s00268-010-0486-4>.
12. Palter VN, MacRae HM, Grantcharov TP. Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg* 2011; **201**: 251–9.
13. Zevin B, Bonrath EM, Aggarwal R, Dedy NJ, Ahmed N, Grantcharov TP. Development, feasibility, validity, and reliability of a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery. *J Am College Surg* 2013; **216**: 955–65.e8.
14. Fink A, Koseoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984; **74**: 979–83.
15. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol* 2003; **56**: 1150–6.
16. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud* 2001; **38**: 195–200.
17. Williams PL, Webb C. The Delphi technique: a methodological discussion. *J Adv Nurs* 1994; **19**: 180–6.