The American Journal of Surgery 217 (2019) 214–221

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

Comparison of faculty versus structured peer-feedback for acquisitions of basic and intermediate-level surgical skills

Guy Sheahan*, Richard Reznick, Don Klinger [1], Leslie Flynn, Boris Zevin

Queen's University, Macklem House, 18 Barrie St., Kingston, Ontario, K7L 3N6, Canada

ARTICLE INFO

Article history:
Received 15 May 2018
Received in revised form
16 June 2018
Accepted 24 June 2018

ABSTRACT

Purpose: Video feedback and faculty feedback has been shown to improve surgical performance; however, consistent access to faculty is challenging. We studied the utility of structured peer-feedback (PF) compared to faculty-feedback (FF) during acquisition of basic and intermediate surgical skills.
Methodology: Two randomized non-inferiority trials were conducted with 1st (n = 30) and 2nd year (n = 29) medical students learning skin-lesion excision and closure (S), and single-layer hand-sewn bowel anastomosis (B), respectively. Five attempts were performed. PF participants used an Objective Structured Assessment of Technical Skills tool to guide feedback. Blinded raters assessed video-recorded performance, time and Integrity of the completed task were also assessed.
Results: For both tasks performance by PF was comparable to FF ($P = 0.111$). Both groups improved significantly: performance ($B{:}P < 0.0001$, $S{:}P = 0.035$), time ($B{:}P = 0.043$, $S{:}P < 0.0001$) and integrity ($B{:}P < 0.0001$, $S{:}P < 0.032$).
Conclusion: Structured peer-feedback is equivalent to faculty-feedback in the acquisition of basic and intermediate surgical skills, giving students freedom to practice independently.

Introduction

Postgraduate surgical education has been required to adapt to several significant societal shifts in recent years. The introduction of restricted working hours, patient safety concerns, economic and efficiency drives from hospitals have all affected the accessibility of residents learning technical skills. Alternative methods of training have subsequently been implemented to augment the traditional learning that occurs in the operating room.[1] In particular, during the early stages of surgical training, surgical boot camps and simulation laboratories have been shown to be effective alternative approaches to teach fundamental surgical skills.[2,3]

Irrespective of the strategies employed, it is the adherence to the evidence-based principals of educational practice that enabled these initiatives to succeed. One of these principles is the provision of timely and high-quality feedback during skill training. Anders Ericsson, through his research on deliberate practice,[4] has been able to demonstrate that provision of high quality feedback during practice is the only way to achieve expert level of performance. Ericsson's research highlighted the importance of feedback[4] provided by an expert; however, in the medical context, consistent access to faculty for feedback remains a challenge.[5]

To overcome some of the challenges with faculty participation in providing feedback, alternatives such as self-assessment strategies or peer-feedback have been studied, with results generally favouring expert feedback over either self or peer approaches.[6–8]

Although these studies have generally compared feedback from trained experts with untrained novice individuals and peers. Studies that have provided more guidance and training to the peer-based feedback groups have in fact demonstrated positive outcomes from peer-based feedback.[9,10] This additional guidance and training provided students with the structured framework necessary to critically review their own, or their peer's performance, thereby gaining insight into the task and guiding improvement in their performance.

The purpose of our study was to build on prior research by further developing the concept of structured guidance. Structured guidance is defined as the reflective practice interventions that are provided to students in order for them to knowledgeably examine their performance of a task. For example, training students to perform an

* Corresponding author.
E-mail addresses: 16ges@queensu.ca (G. Sheahan), richard.reznick@queensu.ca (R. Reznick), dklinger@waikato.ac.nz (D. Klinger), flynnl@queensu.ca (L. Flynn), boris.zevin@kingstonhsc.ca (B. Zevin).
[1] Present Address: University of Waikato, Gate 1 Knighton Road, Private Bag 3105, Hamilton 3240, New Zealand. (Dr Klinger having previously been at Queen's University).

https://doi.org/10.1016/j.amjsurg.2018.06.028
0002-9610/

assessment of the procedure, or to complete a self-reflection. These reflective tasks that are used as part of structured guidance have been shown to improve diagnostic accuracy.[11] But to achieve this outcome of improved diagnostic accuracy, these reflective tasks need to be appropriately introduced into the training curriculum. When poorly implemented, these tasks have been resented by students,[12,13] for example, students that are requested to reflect on their performance, but are not provided with the requisite skills or understand the expectations of the reflection activity.[14] Our primary research objective was to investigate whether students' performance improves with structured peer-feedback as compared to traditional faculty feedback for both basic and intermediate level surgical tasks. Our secondary research objective was to explore if the inclusion of these reflective activities during technical skill practice was perceived to be beneficial by study participants.

Methods

Two prospective randomized non-inferiority controlled trials were conducted with 1st and 2nd year medical students at Queen's University, Canada examining technical skill acquisition under two types of feedback conditions. Structured peer-feedback was the intervention and faculty feedback the control. Study # 1 examined acquisition of intermediate level technical skill - a single layer interrupted handsewn bowel anastomosis using cadaveric porcine small bowel. Study # 2 examined acquisition of basic level technical skill - a skin lesion excision and suture closure using cadaveric porcine skin. The objective of each study was to investigate whether provision of structured peer-feedback resulted in non-inferior technical skills as compared to provision of faculty feedback. All participants in both studies were provided with relevant theory and shown a video demonstration of the task they were required to perform, prior to commencing their individual video recorded practices.

Feedback types

Faculty feedback was based primarily on the three assessment domains used in this study: the technical elements articulated in the objective structured assessment of technical skills (OSATS) assessment tool,[15] the time taken to complete the procedure, and the observed results of the functional assessment of each final product, as described below. The OSATS assessment tool contains a task specific checklist and global rating scale that examines different aspects of technical performance. Faculty members discussed future goals and areas for improvement during the feedback session. Faculty members were trained in how to use the OSATS tool for feedback prior to the commencement of the study. Feedback was individualised and based on direct observation, with a maximum faculty to participant ratio of 1:8.

Participants in the peer-feedback groups provided feedback to each other based on the reflective activities taught to the groups prior to commencing their individual practices. This was achieved by providing structured guidance in how to use the OSATS tool for both assessment and feedback purposes using frame of reference training by assessing a video of a poor performance. Additionally, self-reflection prompts were added into the OSATS documentation to assist in the feedback process, for example, asking assessors to justify why they scored the way they did (Appendix # 1). After participants in the peer-feedback group were satisfied that they understood how to provide peer-feedback using these reflective tools, they commenced practicing the technical skill individually. A self-selected peer from peer-feedback group, reviewed and assessed his/her video using the OSATS tool immediately post completion of the task, reviewed the time taken to complete the procedure, and the observed results of the functional assessment of the final product. Based upon the OSATS score, the time taken and functional assessment results, the peer and study participant collectively developed goals for the participant's subsequent practice attempt.

Study # 1: bowel anastomosis

Participants, inclusion, and exclusion criteria: Participants were recruited from 2nd year medical students at Queen's University who were participating in a summer surgical skills training program. Participants were recruited as a convenience sample via email invitations. Participants were excluded from participating if they could not dedicate enough time to complete a total of 5 bowel anastomosis, and if they were unable to suture or perform an instrument tie for a surgical knot.

Randomization: Closed envelope technique using a sample random number generator.

Description of learned technical skill: Participants were asked to anastomose two segments of transected 25 cm cadaveric porcine small bowel using a single layer interrupted hand-sewn end to end suture technique[15–17] using 3.0 silk or 3.0 Vicryl®. Participants practiced the anastomosis task a minimum of five times over four days, receiving concurrent feedback throughout, with the method of feedback determined by their randomization. Each attempt was video recorded with cameras focussed exclusively on the operative field, de-identifying study participants. At the completion of each anastomosis, a pressure test using a column of water was conducted until leakage occurred (Fig. 1). This test provided additional visual feedback to the participants about the quality of their constructed anastomosis.

Study # 2: skin lesion excision

Participants, inclusion, and exclusion criteria: Participants were recruited from 1st year medical students at Queen's University using the same methodology as Study # 1. Participants were excluded if they were not able to commit to completing 5 attempts, and if they were unable to suture or perform an instrument tie for a surgical knot.

Randomization: Closed envelope technique using a sample random number generator.

Description of learned technical skill: Participants were asked to excise a 3–8 mm diameter skin lesion, painted onto a cadaveric piece of porcine skin, using 3:1 ratio elliptical excision with a minimum 5 mm margin. The skin defect was closed using an interrupted single layer suture technique using 3.0 Prolene® or 3.0

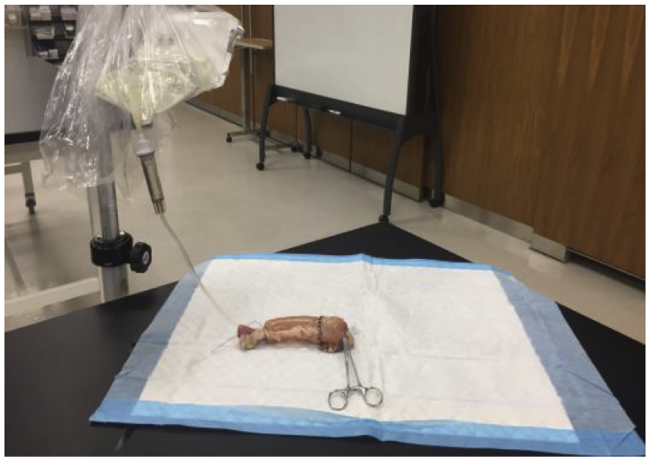

Fig. 1. Anastomotic leak test.

Monocryl®. A functional assessment of the strength of skin closure was performed using a 5 kg accumulated weight displacement test, see Fig. 2.[18] This test provided participants additional visual feedback about the quality of their skin closure. Failure of the functional test was defined as a wound edge displacement of 5 mm or if the suture knots failed or tore through. Participants practiced the task a minimum of five times over four days. All practice attempts were video recorded in a de-identified manner.

Baseline demographic data

Participant age, sex, current year of medical school training, participant expectations and perceptions of learning these skills, and previous technical expertise were obtained at the commencement of the study. Technical expertise was subdivided into suturing (how many suturing opportunities had they had previously, e.g. 3 previous suturing workshops), and exposure to the specific task. Exposure was defined as having seen the task or assisted/performed the task.

Assessment of technical performance post completion of training

Technical performance was assessed across three domains: OSATS assessment of all videos, time to complete the procedure (minutes), and the outcomes of the functional assessments of the completed product as described above. All videos from both control and intervention groups were collated, randomized, using a Microsoft Excel® random number generator, and provided to two trained and blinded reviewers (raters) for assessment using the OSATS tool. Part of rater training included reaching a consensus on interpretation of technical performance prior to assessing the videos of study participants. Raters were asked to submit their own videos of their own performance. One of the videos was then used for the video demonstration and rating. Raters were able to fast forward the video according to their preference.[19] Videos with assessment scores that differed by at least 9 points between raters were re-assessed by both raters to achieve consensus, raters were blinded to their previous scores to prevent forced consensus.

Post study survey examined participants views on the study, the role of feedback, their preferred feedback method, as well as their experience using the reflective activities taught in the study (Appendix # 2). Quantitative questions used a Likert-type scale (1-strongly disagree, 3-neutral and 5-strongly agree). Short answer question responses were collated and examined for emergent themes.

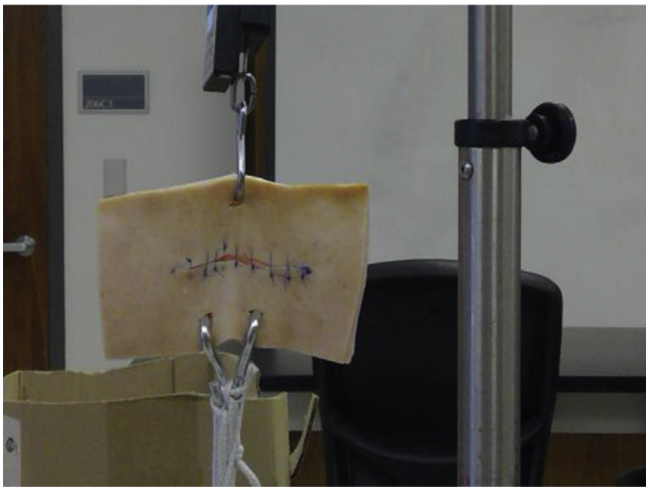

Fig. 2. Skin excision and closure stress test.

Sample size calculation

A sample size of 13 participants in each group was required for each study, based on a non-inferiority power calculation[20] using standard deviation of 3.0,[21] α of 0.05, power of 80%, and δ of 3.0. Both studies were conducted in accordance with university ethics requirements (Queen's University IRB no: 6020826) and were conducted at School of Medicine Simulation Centre.

Statistical analysis

Demographic data was analysed using descriptive statistics. OSATS scores and final product analysis scores are reported as means (standard deviations). Between group comparisons of continuous data was performed using a mixed ANOVA. Within group comparisons were performed using a mixed ANOVA with Bonferroni correction for repeated measures. Inter-rater reliability was calculated using an intraclass correlation two-way random effects model. Survey response comparisons were assessed using a Mann-Whitney *U* test. Data was analysed using SPSS version 24.

Results

Study # 1: bowel anastomosis

Twenty-nine second year medical students (12 males, 17 females, age 25.4(3.5) yrs.) participated in this study. Participants reported previous technical experience of 9.0(5.6) exposures to suturing. Six participants had previously seen a bowel anastomosis and one participant previously assisted in the creation of more than 3 anastomoses. There were no differences between groups in baseline experience ($P = 0.291$). Baseline and fifth attempt assessment scores are shown in Table 1.

Technical skill (OSATS scores)

There was no difference in technical skills at the completion of training between peer-feedback and faculty-feedback groups ($p = 0.111$). Participants in peer-feedback and faculty-feedback groups demonstrated a significant within group improvement in their technical skills across all five attempts ($P < 0.0001$; Fig. 3). Similar results were also obtained for checklist and global rating components of the OSATS tool (Table 2). There was excellent inter-rater agreement for OSATS scores (ICC = 0.844; $P < 0.0001$).

Time to complete the task

There was no statistically significant difference between the peer-feedback and faculty-feedback group in the time required to complete the anastomosis ($P = 0.369$). The time to complete an anastomosis improved over the five attempts for both groups ($P = 0.043$; Fig. 4). Although statistically significant overall, pairwise differences between individual attempts were not significant.

Quality of final product

The quality of the anastomosis, as measured by the water pressure leak test, for peer-feedback and faculty-feedback groups was not significantly different at baseline or at completion of training ($P = 0.474$). The quality of the anastomosis improved over the five attempts for each group ($P < 0.0001$; Fig. 5).

Study # 2: skin excision

Thirty first year medical students (12 males, 18 females, age 24.8(2.4) yrs.) participated in the skin excision study. Participants

Table 1
Bowel anastomosis: performance at baseline and on the fifth attempt.

	Number of participants	Number of attempts	OSATS		Time (mins)		Pressure (cmH20)	
			Baseline	5th attempt	Baseline	5th attempt	Baseline	5th attempt
Faculty	15	5.8 (1.1)	32.2 (6.1)	43.4* (3.2)	39.5 (10.8)	35.6* (5.4)	12.3 (8.1)	23.7* (5.3)
Peer	14	5.2 (0.6)	34.0 (5.2)	43.7* (3.6)	41.2 (12.6)	39.7* (7.8)	15.3 (7.2)	22.8* (6.8)

Starred results are significant (P < 0.05).

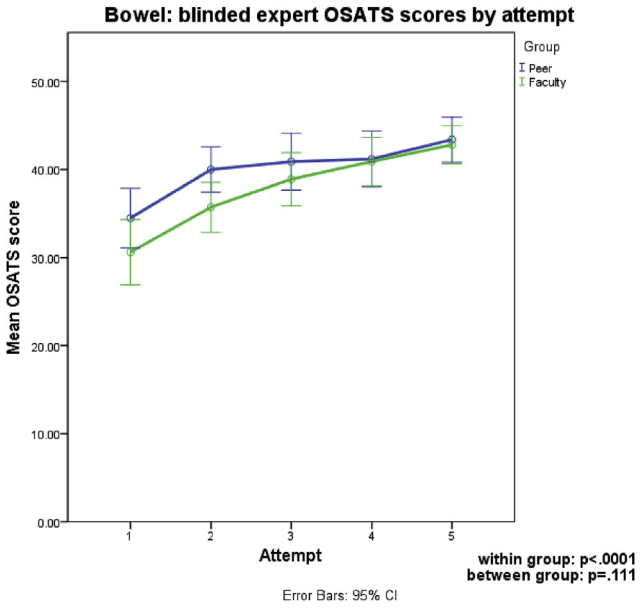

Fig. 3. Bowel: expert OSATS scores by attempt.

reported an average of 5.0(4.2) previous opportunities to practices suturing. Seven participants reporting observing more than 3 excisions or assisting in fewer than 3 skin excisions. There was no difference in baseline level of experience between the groups (P = 0.291). Baseline and post-completion of training assessment scores for peer-feedback and faculty-feedback groups are reported in Table 3.

Technical skill (OSATS scores)

There was no difference in technical skill at the completion of training between peer-feedback and faculty-feedback groups (P = 0.561). Participants in peer-feedback and faculty-feedback groups demonstrated a significant improved in their technical skills across all five attempts (P = 0.035; Fig. 6). Similar results were also obtained when the OSATS was divided into its checklist and global rating components (see Table 4). There was good inter-rater agreement for OSATS scores (ICC = 0.721; P = 0.001).

Time to complete the task

There was no statistically significant difference in duration of time required to complete the skin excision and closure between the peer-feedback and faculty-feedback groups at baseline or at the completion of training (P = 0.487). There was a significant decrease in time required to complete the skin excision and closure over the five attempts for both feedback groups (P < 0.0001; Fig. 7).

Quality of final product

Participants in the peer-feedback and faculty-feedback groups achieved a 100% pass rate on the skin stress test by their third attempt - a significant improvement from the baseline score (P = 0.032; Fig. 8). The quality of final product was not significantly different between the peer-feedback and faculty-feedback groups (P = 0.893)..

Exploration of reflective activities (survey results)

In addition to the Likert-type questions shown in Table 5, participants were asked to describe what they learned from their involvement in the study. These short answer responses were collated and examined for common themes. Three main themes emerged. The first theme was the improvement in technical abilities. Participants (78.8%) reported that this study 'improved (my) suture technique' and that 'I learned how to tie better sutures and how far apart the sutures could be placed while still being effective'. The second theme was the value of practice. Participants (21.2%) reported that they could 'confidently just go at a practice without the stress of trying to make it perfect from the start' and that 'I got to experiment a lot with different techniques (bite sizes, how to grasp tissue)'. The final theme was feedback. Participants (27.3%) reported that they 'learned how to accept feedback and how to effectively reflect on mistakes as a method to improve learning' and 'implementing feedback into manual techniques of suturing'. Of those responses, 75% focussed on the technical aspects of giving or receiving feedback for performance enhancement, only 25% - all in the faculty group, commented on the broader metacognitive aspects of feedback and reflection. There was no significant difference in frequency of responses between the faculty or peer-feedback groups (P = 0.222).

Discussion

The results of our study demonstrate that peer-feedback was equivalent to faculty-feedback in acquisition of basic and intermediate technical skills (skin excision and bowel anastomosis) as assessed by OSATS scores of videotaped reviews of performance, time to completion of the task and quality of the final product. Furthermore, performance of participants in both the peer-feedback and faculty-feedback groups improved significantly from baseline over the five attempts.

Our study utilised a process of deliberate practice for acquisition of technical skills. Deliberate practice, as identified by Ericsson, is

Table 2
Bowel anastomosis: OSATS scores subdivided into checklist and global rating scale.

	OSATS 1st Attempt			OSATS 5th Attempt					
	Checklist	Global Rating Scale	Total	Checklist	Within group	Global Rating Scale	Within group	Total	Within group
Faculty	15.1 (2.6)	17.1 (4.0)	32.2 (6.1)	17.6 (1.0)	P = 0.003	25.6 (3.1)	P < 0.001	43.3 (3.2)	P = 0.001
Peer	16.0 (2.3)	18.0 (3.1)	34.0 (5.2)	17.9 (1.0)	P = 0.014	25.8 (3.2)	P = 0.001	43.7 (3.6)	P < 0.001
Between groups	P = 0.563	P = 0.371	P = 0.433	P = 0.977		P = 0.970		P = 0.820	

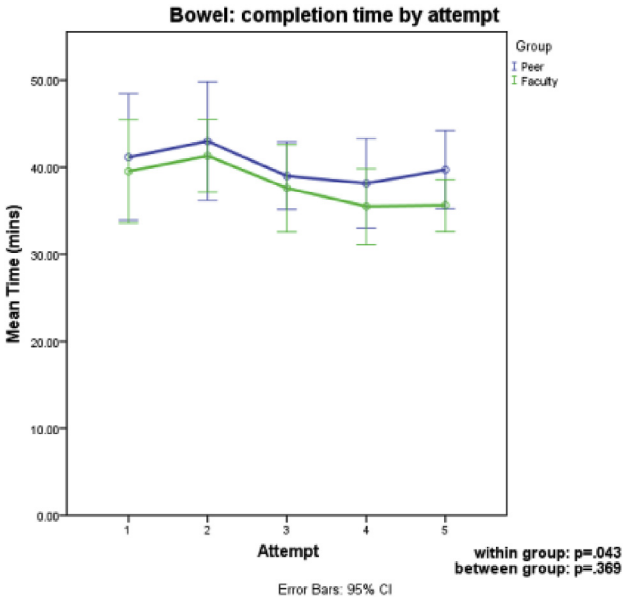

Fig. 4. Bowel: completion time by attempt.

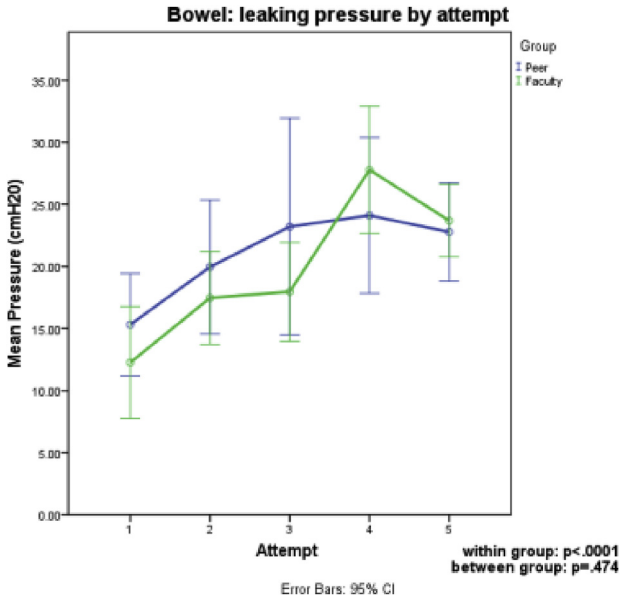

Fig. 5. Bowel: anastomosis leak pressure by attempt.

based on the concept that it is the application of relevant feedback, combined with opportunities for repetition, that is necessary to achieve expertise.[4] In order to provide relevant feedback, it is necessary to first adequately define the task itself. We deliberately chose three assessment criteria to define each task. These criteria were chosen because they could be easily understood by the participants, and could be used to set individual learning goals.[8]

We deliberately chose to use the OSATS assessment tool because the checklist component acted as a prompt for the peer-feedback group, allowing participants to gain familiarity with the steps of the procedure. The use of video recording allowed participants in the peer-feedback group to practice independently and still obtain relevant peer-feedback through video review. Although it is difficult to separate the educational effects of assessing the video footage from the other sources of feedback, the improvement in OSATS scores does suggest that using video footage for feedback was beneficial. This is consistent with other studies that have demonstrated an educational benefit of using video review.[9,22] Reviewing video footage, however, takes time. Editing video footage for the purpose of assessment[23] is time intensive and may exclude critical issues; however, it does provide opportunities for focused feedback and reduces the time required for video review. Faculty did not use the recorded video for feedback purposes, as the faculty group was the control group for our study and as such, participants were taught in the traditional manner using contemporaneous feedback.

Improvement in technical skills was more pronounced in the bowel anastomosis study than in the skin excision and closure study. This result could have been influenced by the fidelity of the skin model, as participants found the thick porcine skin difficult to manipulate, especially when it dried out during the long duration of practice, thereby making it difficult to adequately demonstrate careful tissue handling. Additionally, the use of trained raters for video assessment, rather than in-situ real-time assessment, may have influenced our results. Nicke et al.[24] noted that video assessment combined with procedural time may be more comparable to real-time scoring, due to the smoothing effect of reviewing the video at higher speed.

The duration of time required to perform each procedure was used as a form of formative feedback by participants in each group. We did not set specific time goals, however, participants often set personal time goals during the study. For the anastomosis study the overall improvement in time was statistically significant, but the clinical significance (2 min) is questionable, as the improvement between individual attempts was small. It is likely that the participants were pursuing quality improvement of the anastomosis over improvement in time to complete the task. In fact, our results demonstrate that for more complicated tasks time is a poor predictor of technical skill or the quality of the final product. Participants in Study # 1 took longer to perform their second attempt than their first attempt, likely because they realised that poor technique was correlated with anastomotic failure during the leak test.

Our study assessed the quality of the final product – adequacy of bowel anastomosis and skin closure – which provided participants with visual feedback about the adequacy of their technique, information that was then used for self and peer-assessment. For example, leaks identified at the mesenteric border of the anastomosis became the focus of subsequent practice. Participants in the peer-feedback group felt that they did not possess the required technical expertise to provide high quality feedback; however, they

Table 3
Skin excision: performance at baseline and on the fifth attempt.

	Number of participants	Number of attempts	OSATS		Time (mins)		Force Pass Rate (5 kg)	
			Baseline	5th attempt	Baseline	5th attempt	Baseline	5th attempt
Faculty	15	6.1 (0.9)	36.5 (7.2)	41.9* (5.3)	53.6 (13.8)	31.0* (8.9)	0.92 (0.3)	1.0* (0)
Peer	15	5.5 (0.6)	36.0 (7.1)	39.2* (8.7)	58.6 (16.3)	32.5* (5.9)	0.75 (.45)	1.0* (0)

Starred results are significant ($P < 0.05$).

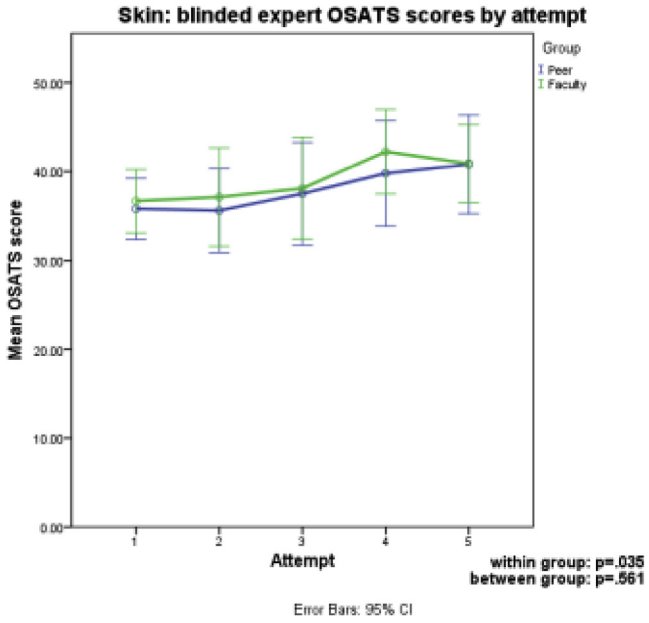

Fig. 6. Skin: expert OSATS by attempt.

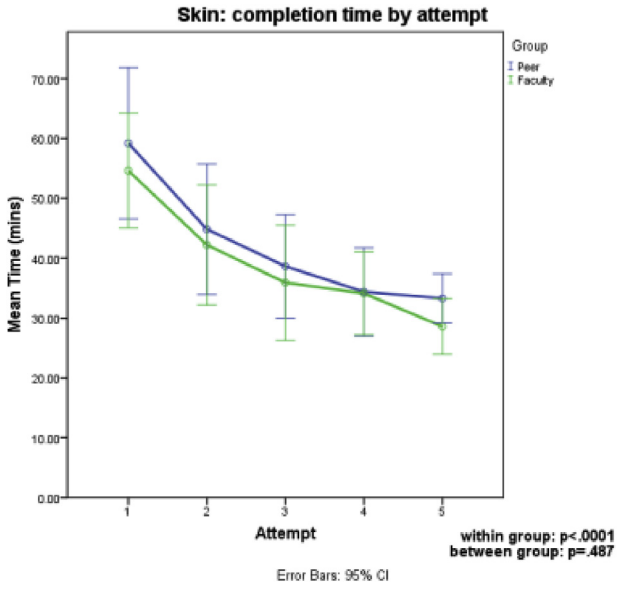

Fig. 7. Skin: completion time by attempt.

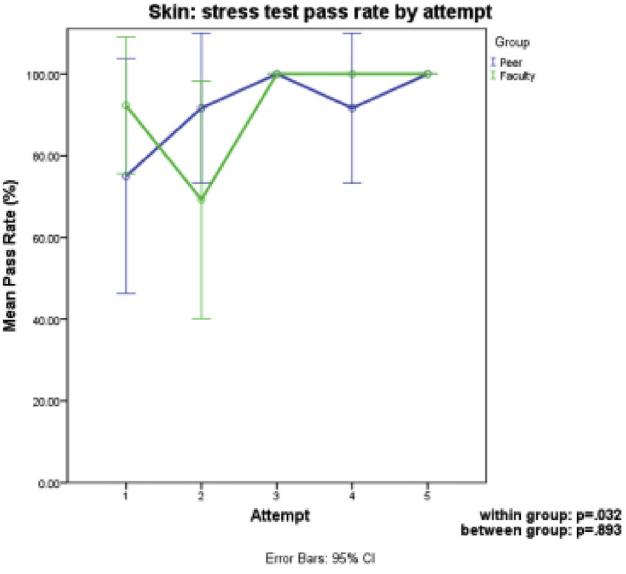

Fig. 8. Skin: stress test pass rate by attempt.

were able to provide highly relevant common-sense suggestions after seeing the quality of the final product. For example, after sutures tore through on a skin stress test, a suggestion was made to place the sutures wider on the next attempt.

Deliberate practice is dependent of provision of timely and accurate feedback. Participants in the peer-feedback group expressed concerns that their lack of expertise would put them at a disadvantage as compared to faculty-feedback group. This concern was highlighted in our post-intervention survey results. Some peer group participants, in fact, came to the investigators at the completion of the study asking for faculty feedback thinking that they received inadequate feedback from their peers. This subjective concern, however, was not supported by our objective results. Participants in both groups in both studies improved to the same degree regardless of the type of feedback that they received.

Despite the subjective fear about lack of expertise of participants in the peer-feedback group, students in this group were able to learn the skill without faculty support and felt able to provide useful feedback to their peers. This highlights that self-directed learning, which is heavily reliant on student's internal motivation, can overcome this fear of inexperience, if appropriate structured support is provided. It is interesting to note that the second-year medical students were slightly more confident in their ability to provide feedback and learn independently. This confidence may have created a small power differential between peer-feedback participants on giving and receiving feedback, but baseline experience demonstrates that as a group they were comparable.

The results of the reflective activities survey demonstrate that some participants in each group, perceived that reflecting on one's actions was a valuable, learnable skill that could be used to improve another's performance as well as his/her own. We did not actively promote the value of self-reflection in each study; however, we hypothesized that by integrating these reflective activities into our study, we could naturally increase the awareness of participants about the importance of self-reflection. Only a small subset of participants (9.1%) reported that self-review or reflection had benefits beyond direct performance improvement, but whether this translates to changes in behaviour beyond the study intervention was outside the scope of this study.

Table 4
Skin excision: Checklist and Global Rating Scale components of the OSATS scores.

	OSATS 1st Attempt			OSATS 5th Attempt					
	Checklist	Global Rating Scale	Total	Checklist	Within group	Global Rating Scale	Within group	Total	Within group
Faculty	17.5 (3.3)	19.0 (4.6)	36.5 (7.2)	20.0 (1.4)	P = 0.030	21.9 (4.8)	P = 0.192	41.9 (5.3)	P = 0.100
Peer	17.6 (3.8)	18.4 (4.3)	36.0 (7.1)	18.8 (3.3)	P = 1.000	20.3 (6.5)	P = 0.618	39.2 (8.7)	P = 0.718
Between groups	P = 0.940	P = 0.719	P = 0.852	P = 0.231		P = 0.499		P = 0.343	

Table 5
Responses for Likert survey questions.

	Skill	Bowel anastomosis		Skin excision and suture closure	
Pre-study survey	How would you rate your current ability to do the skill? (0–10)	1.4 (1.7)	P[a] = 0.134	3.8 (2.3)	P[a] = 0.525
	How would you rate your confidence to learn to the skill? (0–10)	5.8 (3.0)	P[a] = 0.434	7.0 (2.8)	P[a] = 0.080
Post-study survey	Did you think this study was successful in teaching you the skill? (1–5)	4.5 (0.5)	P[a] = 0.601	4.5 (0.5)	P[a] = 0.442
	How comfortable were you learning the skill without faculty to guide you? (1–5)	3.6 (0.8)	N/A	3.3 (0.8)	N/A
	Would you have preferred some faculty feedback in addition to the peer feedback? (1–5)	3.4 (1.0)	N/A	3.5 (0.7)	N/A
	How comfortable were you assessing your work with a peer? (1–5)	3.4 (0.8)	N/A	3.2 (1.2)	N/A

[a] Mann-Whitney *U* test between faculty and peer-feedback groups.

Our study has several limitations. First, study participants were not trained to proficiency, so it is difficult to ascertain whether there are limitations in peer-feedback in acquisition of technical skills at higher levels of performance. We, however, believe that structured peer-feedback can be used in the early stages of learning allowing independent practice, prior to re-engagement with faculty for honing of skills prior to summative assessment. Second, the control and intervention groups were not spatially separated during the conduct of the study. It is possible that sharing of feedback suggestions may have occurred between participants in different groups. This does not necessarily detract from the results, as sharing of feedback suggestions would still be considered peer-based feedback. Third, the faculty-feedback group did not receive one-on-one feedback as the maximum faculty to student ratio was 1:8. This ratio was consistent with other laboratory skill-based activities; however, participants who were having more difficulties may have received more feedback than others. Participants in the peer-feedback group did receive one-on-one feedback with another peer group participant. It is possible that participants in the peer-feedback group received greater quantity of feedback; however, we were not able to quantify this in our study. Providing meaningful feedback is a time intensive process, the quality of feedback for both groups may have suffered due to this constraint. In the faculty group, faculty members had to be shared amongst the participants. In the peer group wanted, peers wanted to return to their own practice. Whether the peer group viewed the assessment and feedback of another's work as beneficial to their own is not known. Finally, the video recordings of practice sessions by the participants in the faculty-feedback participants were not utilised for feedback purposes in this group. Video recordings may have highlighted other areas for improvements not witnessed by the faculty, which could have been addressed at that stage.

We propose several areas of additional inquiry based on our findings. From the technical skill point of view, it would be valuable to see how this approach applies to advanced surgical skills, and whether more experienced participants are more comfortable giving and receiving structured peer-feedback. It would be valuable to examine the effect of incorporating peer-feedback into a post-graduate training curriculum to examine if it may be used to augment or replace faculty-feedback during acquisition of advanced surgical skills. Finally, given the importance of self-reflection as an essential life-long learning skill, it would be valuable to know whether the structured supports provided in this study were internalised by participants, and then applied outside of the study context.

Conclusion

If you provide students with the structural skills necessary to assess performance and provide feedback, then the peer-feedback provided is comparable to faculty feedback for the acquisition of both a basic and intermediate surgical skill. This means that students can be given the freedom to practice independently, without needing to have a faculty member present at all stages of learning a skill. Additionally, if self-review is built in as a natural part of the activity, some students will recognise the importance of these essential life-long skills.

Funding

This work was supported through a scholarship from the Royal Australasian College of Surgeons.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix 1. Reflective prompts built into OSATS

Global Assessment						
1	Respect for tissue	1	2	3	4	5
Why did you give this score?						
2	Time and motion	1	2	3	4	5
Why?						
3	Instrument handling	1	2	3	4	5
Why?						
4	Knowledge of Instruments	1	2	3	4	5
Why?						
5	Flow of Operation	1	2	3	4	5
Why?						
6	Knowledge of specific procedure	1	2	3	4	5
Why?						
Total Score: checklist + Global Scale:			Goals for next attempt:			
Other points to consider reviewing:						
1	Ergonomics of body position: correct height, neck position					
2	Try focussing on one element					
3	Set a time goal					
4	Have you compared your performance with a video?					

Appendix 2. Post study survey questions

Did you think this study/education was successful in teaching you to do the skill?	No	Not really	Neutral	Somewhat	Yes
How comfortable were you learning to the procedure without an expert to guide you?	Very un-comfortable	Un-comfortable	Neutral	Comfortable	Very comfortable
Did the teaching session accommodate your learning style?	No	Not really	Neutral	Somewhat	Yes
Biggest strengths of the teaching session/study	Open response				
Biggest weaknesses of the teaching session/study	Open response				
What did you learn from this activity/study?	Open response				
Any other comments?	Open response				
Which group: peer or expert, did you want to be in for learning how to do the procedure? Why?	Open response				
Would you have preferred to have some expert feedback in addition to the peer-feedback?	No Attempts	Some attempts	Half the attempts	Most attempts	All attempts
If you wanted expert feedback in addition to peer-feedback, when would you like it? - choose all that apply	All the way through	Beginning (e.g. first attempt)	End (e.g. last attempt)	Middle	First half Last half
How comfortable were you reviewing/scoring your work with a peer?	Very un-comfortable	Un-comfortable	Neutral	Comfortable	Very comfortable

References

1. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. Effect of practice on standardised learning outcomes in simulation-based medical education. *Med Educ*. 2006;40(8):792–797. https://doi.org/10.1111/j.1365-2929.2006.02528.x.
2. Fernandez GL, Page DW, Coe NP, et al. Boot cAMP: educational outcomes after 4 successive years of preparatory simulation-based training at onset of internship. *J Surg Educ*. 2012;69(2):242–248. https://doi.org/10.1016/j.jsurg.2011.08.007.
3. Levine AI, DeMaria S, Schwartz AD, Sim AJ, eds. *The Comprehensive Textbook of Healthcare Simulation*. New York, NY: Springer New York; 2013. https://doi.org/10.1007/978-1-4614-5993-4.
4. Anders Ericsson K. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med*. 2008;15(11):988–994. https://doi.org/10.1111/j.1553-2712.2008.00227.x.
5. Zevin B, Sheahan G, Ashamalla S, Dedy NJ, Jalink D, Grantcharov T. *Implementation and Evaluation of a Comprehensive Proficiency-based Upper GI Curriculum: A Multi-institutional Canadian Experience*. 2018. Chicago, IL.
6. Jamshidi R, LaMasters T, Eisenberg D, Duh Q-Y, Curet M. Video self-assessment augments development of videoscopic suturing skill. *J Am Coll Surg*. 2009;209(5):622–625. https://doi.org/10.1016/j.jamcollsurg.2009.07.024.
7. Nesbitt CI, Phillips AW, Searle RF, Stansby G. Randomized trial to assess the effect of supervised and unsupervised video feedback on teaching practical skills. *J Surg Educ*. 2015;72(4):697–703. https://doi.org/10.1016/j.jsurg.2014.12.013.
8. Porte MC, Xeroulis G, Reznick RK, Dubrowski A. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *Am J Surg*. 2007;193(1):105–110. https://doi.org/10.1016/j.amjsurg.2006.03.016.
9. Vaughn CJ, Kim E, O'Sullivan P, et al. Peer video review and feedback improve performance in basic surgical skills. *Am J Surg*. 2016;211(2):355–360. https://doi.org/10.1016/j.amjsurg.2015.08.034.
10. Hoogenes J, Mironova P, Safir O, et al. Student-led learning: a new teaching paradigm for surgical skills. *Am J Surg*. 2015;209(1):107–114. https://doi.org/10.1016/j.amjsurg.2014.08.037.
11. Mann KV. Reflection's role in learning: increasing engagement and deepening participation. *Perspect Med Educ*. 2016;5(5):259–261. https://doi.org/10.1007/s40037-016-0296-y.
12. Larrivee B. Development of a tool to assess teachers' level of reflective practice. *Reflective Pract*. 2008;9(3):341–360.
13. Anderson PAM. Giving feedback on clinical skills: are we starving our young? *J Grad Med Educ*. 2012;4(2):154–158. https://doi.org/10.4300/JGME-D-11-000295.1.
14. Russell T. Can reflective practice be taught? *Reflective Pract*. 2005;6(2):199–204. https://doi.org/10.1080/1462394050010583.
15. Martin J, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84:273–278.
16. Egle JP, Malladi SVS, Gopinath N, Mittal VK. Simulation training improves resident performance in hand-sewn vascular and bowel anastomoses. *J Surg Educ*. 2015;72(2):291–296. https://doi.org/10.1016/j.jsurg.2014.09.005.
17. Shah J, Munz Y, Manson J, Moorthy K, Darzi A. Objective assessment of small bowel anastomosis skill in trainee general Surgeons and urologists. *World J Surg*. 2006;30(2):248–251. https://doi.org/10.1007/s00268-005-0074-1.
18. Westlake Erica. Breaking strength of suture technique. *Academy of Notre Dame de Namur*; 2012. http://www.ndapa.org/uploaded/academics/docs/Science_Fair/Erica_Westlakes_paper[1].pdf. Accessed June 8, 2017.
19. Dath D, Regehr G, Birch D, et al. Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. *Surg Endosc*. 2004;18(12):1800–1804. https://doi.org/10.1007/s00464-003-8157-2.
20. Sealed Envelope | Power calculator for continuous outcome non-inferiority trial. https://www.sealedenvelope.com/power/continuous-noninferior/. Accessed April 15, 2017.
21. Olson TP, Becker YT, McDonald R, Gould J. A simulation-based curriculum can Be used to teach open intestinal anastomosis. *J Surg Res*. 2012;172(1):53–58. https://doi.org/10.1016/j.jss.2010.08.009.
22. Soucisse ML, Boulva K, Sideris L, Drolet P, Morin M, Dubé P. Video coaching as an efficient teaching method for surgical residents—a randomized controlled trial. *J Surg Educ*. 2017;74(2):365–371. https://doi.org/10.1016/j.jsurg.2016.09.002.
23. Datta V, Bann S, Mandalia M, Darzi A. The surgical efficiency score: a feasible, reliable, and valid method of skills assessment. *Am J Surg*. 2006;192(3):372–378. https://doi.org/10.1016/j.amjsurg.2006.06.001.
24. 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 Basel*. 2016;57(1-2):1–9. https://doi.org/10.1159/000444449.