Establishing Reliability of A Coding Scheme for the Qualitative Assessment of Narrative Feedback through Society for Improving Medical Professional Learning

Ana Isabel Jacinto MD[1], Tandis Soltani, MD[2], Avery Wilson, BS[3], Jessica Tsang, BS[3], Vitaliy Popov, PhD[3]

[1]University of California, Davis, Department of Surgery
[2]Loma Linda University Health, Department of Surgery
[3]University of Michigan, Department of Learning Health Sciences

Please contact Ana Jacinto at aljacinto@ucdavis.edu with any questions

Introduction

The American Board of Surgery has selected the Society for Improving Medical Professional Learning (SIMPL) application as a platform for providing feedback to general surgery residents.[1]

Feedback historically has been unstructured and untimed, with most feedback lacking proper structure and/or actionable strategies.

A coding scheme to assess the quality of feedback does not exist.

Therefore, we aimed to create a procedure-specific coding scheme and establish its inter-rater reliability for assessment of the quality of narrative feedback.

Methods

Narrative feedback on laparoscopic cholecystectomies performed from 2015-2021 was extracted from the SIMPL platform. It was limited to PGY2-3 surgical residents. Data was deidentified.

After an initial review of 100 cases, a coding sheet with 7 themes and 3 levels was created against which the narratives were graded.

Three coders underwent rounds of social moderation to understand the themes and levels to more accurately code the feedback.

Table 1: Seven Themes of Narrative Feedback

	Title	Description	Example
1	Resident Applicability	Feedback is pertinent to the case that was performed	“Resident performed well and it was an extremely difficult laparoscopic cholecystectomy with cholangiogram.”
2	Operative Steps	Comments on specific step: port placement, obtaining critical view of safety, isolation of cystic artery and/or duct	“All ports were placed safely and with appropriate technique.” “Good job opening up the peritoneum and exposing your cystic plate.”
3	Technical Skills	Evaluates instrument handling, respect for tissue, use of left hand, creating tension and counter-tension	“You showed good instrument control” “I think your retraction with your left hand on the gallbladder was effective.”
4	Decision Making	Comments on clinical judgement, understanding of the dissection planes, progressing the case forward	“I thought you did a much better job of seeing the plane.” “He has good judgment and he approached the gallbladder in an organized fashion.”
5	Efficiency	Evaluates pace and/or comments on efficiency	“Good efficiency in progressing through the case.”
6	Communication /Coordination	Evaluating communication with other staff, taking feedback, teaching skills	“He was easy to instruct.” “He did a really good job teaching and taking his junior through the case.”
7	Stress Management	Comments on resident’s emotions during stressful part of case	The gallbladder was difficult to extract through the epigastric port and she did exhibit some impatience.”

Methods

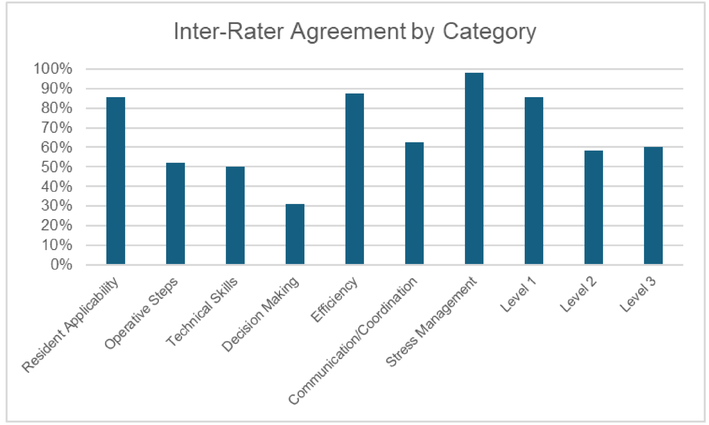
Table 2: Three Levels of Narrative Feedback (derived from Gabelica et. al 2014)[2]

Level	Title	Description	Example
1	Evaluating performance	Assessment of the situation and/or performance. Comments on what went well and what did not. This is strictly looking at the past.	“She did well, created her critical view, I never took control of the case.” “Excellent tissue handling and good spatial awareness.”
2	Looking for alternatives	Suggestions or discussions in what the trainee could have done differently. Identifying the gap in the performance. This is still looking at the past performance.	“I would have preferred you to use the Maryland as more of a dissector as opposed to ripping the peri-areolar tissue around the structures.” “I would have started maybe a little lower on the gallbladder.”
3	Looking ahead	Statements depicting a new direction to take, how to do things differently. This is a concrete and actionable strategy for the resident to take moving forward.	“I think we should really work on two-handed laparoscopic skills.” “Your next step is to work on self-directed movements and speed of the operation.”

For the inter-reliability analysis, 30 new cases were randomly selected from the dataset and coded by three separate coders.
Inter-reliability was evaluated using Cohen's kappa coefficient for paired comparisons between coders and Fleiss' kappa for overall agreement across all three coders.

Results

Figure 1: Perfect Inter-Rater Agreements by Category



Overall perfect agreement across three raters and categories was 67.1%
Operative steps (52.1%), technical skills (50%), decision making (31%) had the lowest agreement. Efficiency (87.5%) and stress management (97.9%) had the highest.

Results

Figure 2: Pairwise Agreements by Category

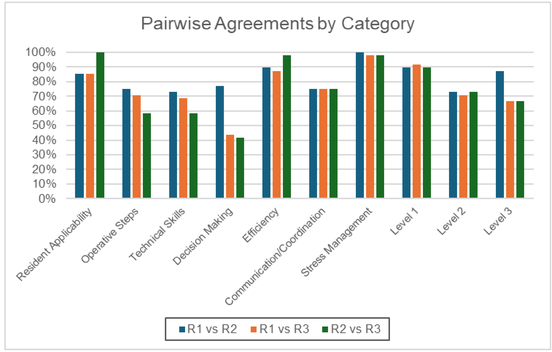
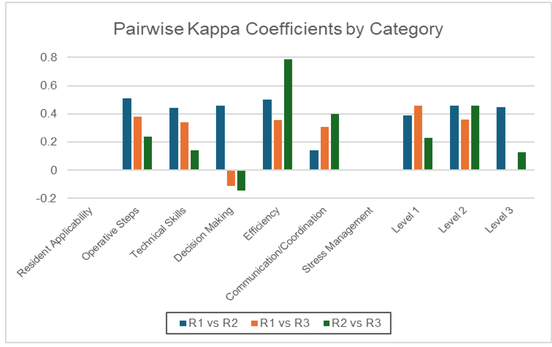


Figure 3: Pairwise Kappa Coefficients by Category



Legend
- R1: Rater 1
- R2: Rater 2
- R3: Rater 3

Overall pairwise agreement (all categories combined):
- R1 vs R2: 82.5%
- R1 vs R3: 75.8%
- R2 vs R3: 75.8%

Two categories had the highest pairwise agreement at 100%. This was seen for resident applicability between R2 and R3 and stress management between R1 and R2. The lowest pairwise agreement was seen with the category of decision making between R2 and R3 (41.7%).

Cohen’s kappa coefficient[3]:
- 0-0.20 = no agreement
- 0.21-0.39 = minimal agreement
- 0.40-0.59 = weak agreement
- 0.60-0.79 = moderate agreement
- 0.80-0.90 = strong agreement
- >0.90 = near perfect agreement

For resident applicability, insufficient data was obtained between R2 and R3 resulting in no Kappa coefficient and no graph. The same was seen for stress management between R1 and R2. The highest inter-rater reliability was seen for efficiency between R2 and R3. K = 0.789 (moderate).

Conclusions

This is a reliable coding scheme that can be used to assess the quality of narrative operative feedback. With further refinement, it can be used to evaluate a larger dataset, not just limited to PGY2 and PGY3 residents, and across multiple institutions. This coding scheme has the potential to reveal patterns and systemic trends in feedback that can inform future initiatives in improving narrative feedback for surgery residents.

References

1. Zwischenberger JB, Hatchett SP, Rahman NA, Chadha P, Sebok-Syer SS, George BC. Implementing Workplace-Based Assessments at Scale: The SIMPL Approach. Ann Surg Open. 2023 Nov 27;4(4):e353. doi: 10.1097/AS9.0000000000000353. PMID: 38144481; PMCID: PMC10735067.
2. Gabelica, Catherine & Van den Bossche, Piet & R.Segers, Mien & Gijselaers, Wim. Dynamics of Team Reflexivity after Feedback. Frontline Learning Research. 2014 June 18;2(3). doi: 10.14786/flr.v2i2.79
3. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82. PMID: 23092060; PMCID: PMC3900052.