Quality of dictated feedback associated with SIMPL operative assessments of pediatric surgical trainees

Benjamin Zendejas[*], Ayca Toprak, Amanda W. Harrington, Craig W. Lillehei, Biren P. Modi

Department of Surgery, Boston Children's Hospital, Boston, MA, USA

ARTICLE INFO

ABSTRACT

Background: SIMPL is a workplace-based operative performance assessment tool which allows for dictated feedback (DF). To better understand the value of DF, we sought to characterize the type and quality of DF generated during SIMPL evaluations.
Methods: Thematic analysis of DF from SIMPL assessments between June 2017 and December 2018 at a single pediatric surgery fellowship program was performed. Comments were categorized as specific, encouraging or corrective. Categories were combined to determine DF quality as effective, mediocre or ineffective.
Results: Of 781 SIMPL assessments (21 faculty, 5 trainees), 451 (57%) had DF. Most comments were encouraging (93%) and specific (65%). Only 21% were corrective, 17% had entrustment features, and 8% had an explicit learning plan. Feedback quality was deemed mediocre (45%), ineffective (33%) and effective (21%).
Conclusion: SIMPL dictated feedback was mostly encouraging and specific. To improve quality, feedback should incorporate learning plans as well as corrective and entrustment features.

© 2020 Elsevier Inc. All rights reserved.

Introduction

Feedback has proven to improve performance across several disciplines and settings, and is considered of utmost importance for learning.[1–4] The real question is not whether feedback works but under what circumstances it works best. Unfortunately, feedback on performance of surgical trainees appears to be highly variable in quality and quantity.[5,6] Several challenges to deliver feedback exist, including time constraints and fear of being labeled as intimidating or harassing.[7] Feedback is often given to trainees along with summative assessments in the form of end of rotation evaluations that rate the trainee's performance within certain core competencies, milestones and/or entrustable professional activities.[8,9] Recently, work-place based assessments have gained popularity as they allow for more opportunities to provide timely formative feedback in a clinical environment with direct observations of clinical tasks or procedures.[10,11]

SIMPL (System for Improving and Measuring Procedural Learning) is a novel workplace-based operative performance assessment tool which allows case by case assessment of trainees using three scales (autonomy, performance, case complexity) and allows for dictated narrative type feedback.[12–14] SIMPL as an assessment platform has demonstrated robust validity evidence from multiple sources,[15,16] and has shown to significantly increase the quantity of feedback that surgical trainees receive.[17] In addition, trainees have rated dictated feedback as the most valuable item in SIMPL evaluations.[17] However, much less is known about the characteristics and quality of such dictated feedback component.

To better understand the value and effectiveness of dictated feedback, we sought to characterize the type and quality of feedback generated by SIMPL evaluations. We hypothesized that dictated feedback quality would be of variable quality and predominantly encouraging.

Methods

We conducted a qualitative analysis of narrative dictated feedback as it pertains to operative performance of pediatric surgery fellows in a single pediatric surgical training program. Exemption status from our Institutional Review Board was obtained.

SIMPL, a smartphone-based application, is a workplace-based assessment that was developed and is supported by the

---

* Corresponding author. Department of Surgery, Boston Children's Hospital, Harvard Medical School, 300 Longwood Ave., Boston, 02115, MA, USA.
E-mail address: benjamin.zendejas@childrens.harvard.edu (B. Zendejas).

Table 1
Dictated Feedback Features.

Definition	Examples	Counterexamples
SPECIFIC		
Provides context to instructions or statements (i.e. How? Why? When?)	"able to follow tissue planes while removing the mass from the iliac vessels"	"Excellent inter-personal skills... approachable by staff and students."
Detailed statements or qualities belonging or relating uniquely to a particular subject, anatomy, individual, situation, relation, effect, etc.	"made progress on instrument tying in narrow, deep incisions and in recognizing tissue planes"	"Appropriately respectful with patients, well-liked by staff and co-workers."
Clearly defined or identified; free from ambiguity	"capable of a very nice subcuticular closure in a timely manner"	"Always respectful of others, interacts well with the team, is confident but not arrogant, level-headed, calm, street smart (i.e. gets things done), proactive."
Precise and clear in making statements or issuing instructions	"able to…palm instruments to better maneuver in tight spaces"	"Overall an excellent trainee"
Includes all moderate and specific comments when grading degree of specificity		"She was always pleasant, easy-going, calm."
ENCOURAGING		
Giving support, courage or confidence	"Very nice"	"Adequate"
Recognizing effort	"Good/excellent/nice job/work"	"OK"
Positive statement about trainee's character, attitude, skill or technique	"Good outcome"	"Appropriate skill level"
Positive and giving hope/promise for future success; complementary	"Performs ___ well"	"Appropriate for PGY"
	"Well done"	"Appropriate with patients"
	"[Trainee] is/was good at ___"	"No issues"
		"No concerns"
CORRECTIVE		
Designed to correct, rectify, or counteract something harmful or undesirable; remedial	"I would like for you to speed it up some"	"[Trainee] is left-handed which puts him at a disadvantage when trying to use instruments designed for right-handed surgeons"
Statement aimed to remedy or cause improvement	"Next time I would like you to do/not do……"	"[Trainee] works well with the entire OR team"
	"During this case you seemed to struggle with the dissection, I think it is because you weren't using your left hand as well to create counter-tension, next time make sure you pay attention to that"	Remember to …… (clinical or technical pearl)
		The key to this procedure is….(clinical or technical pearl)
ENTRUSTMENT LEVEL		
Defines learners' level of independence or skill relative to expected level of performance	"Practice ready" "Independent"	Comments that describe degree of support needed without explicit comment on level of performance
Must make an explicit comment about resident's performance related to expected level or degree of autonomy	"Getting close to independence"	"overall excellent job, technically confident, no help from me."
	Above/at/bellow level of training	"overall case went well, there was minimal input and none that needed correcting. We had one problem where the clip slipped and there was leakage of some stones, but all the stuff taken correctly and safely"
	"Today you clearly demonstrated that you are practice ready, I encourage you to bring a resident in next time so that you can take them through the case and practice that aspect"	
LEARNING PLAN		
Clearly identifies a learning goal	"We need to work on your efficiency with laparoscopic knot tying, let's get together at the SIM center and work on getting smoother at your movements and knowing how to anticipate which is your next move to set up your knot."	"Today things did not go smoothly, you need to figure out what's going on and make a change"
Provides a recommendation on how to address learning goal		

Procedural Learning and Safety Collaborative (PLSC), a not-for-profit educational research collaborative. SIMPL asks faculty raters to assess directly observed operative performance within 72 h of the completion of a case. It includes three scale-based questions asking about operative performance, trainee autonomy and patient specific case complexity.[12] Additionally, faculty can dictate formative feedback for the trainee about that specific case. It was this latter data that were analyzed for the present study.

SIMPL was implemented in our pediatric surgical fellowship training program in June 2017. Prior to implementation, all users (faculty and trainees) underwent an hour long standardized frame of reference rater training.[18] Rater training focused mostly on rater calibration of operative autonomy and performance scales. Though dictated feedback was encouraged, it remained optional and no specific instruction on how to best give dictated feedback was given. All operative cases performed by surgical trainees were eligible for a SIMPL evaluation. SIMPL dictations that were recorded from June 2017 to December 2018 were transcribed using Google Cloud Speech-to-Text, (Mountain View, California). Transcriptions were de-identified and transcription errors corrected. The unit of analysis was the entire comment. Transcriptions were reviewed by two independent surgeon raters. Thematic analysis was undertaken to identify feedback themes and establish thematic frequencies. Feedback themes were categorized as primary or secondary based on their relationship with the main message of the comment. Comments not related to operative performance were excluded. Feedback narratives were screened to identify and categorize comments as having encouraging, specific, and/or corrective feedback features. Comments that commended a trainee's performance were labeled as encouraging, contrasted with corrective feedback that suggested a need for change. Specific feedback contained comments on specific skills, as opposed to vague observations. These feedback features were chosen based on the feedback classification system proposed by Shaugness et al.[19] In addition, we modified the classification by evaluating dictated narratives for the presence or absence of explicit entrustment or learning plan statements (Table 1). We define entrustment as actions by the faculty that impart trust and responsibility for patient care to
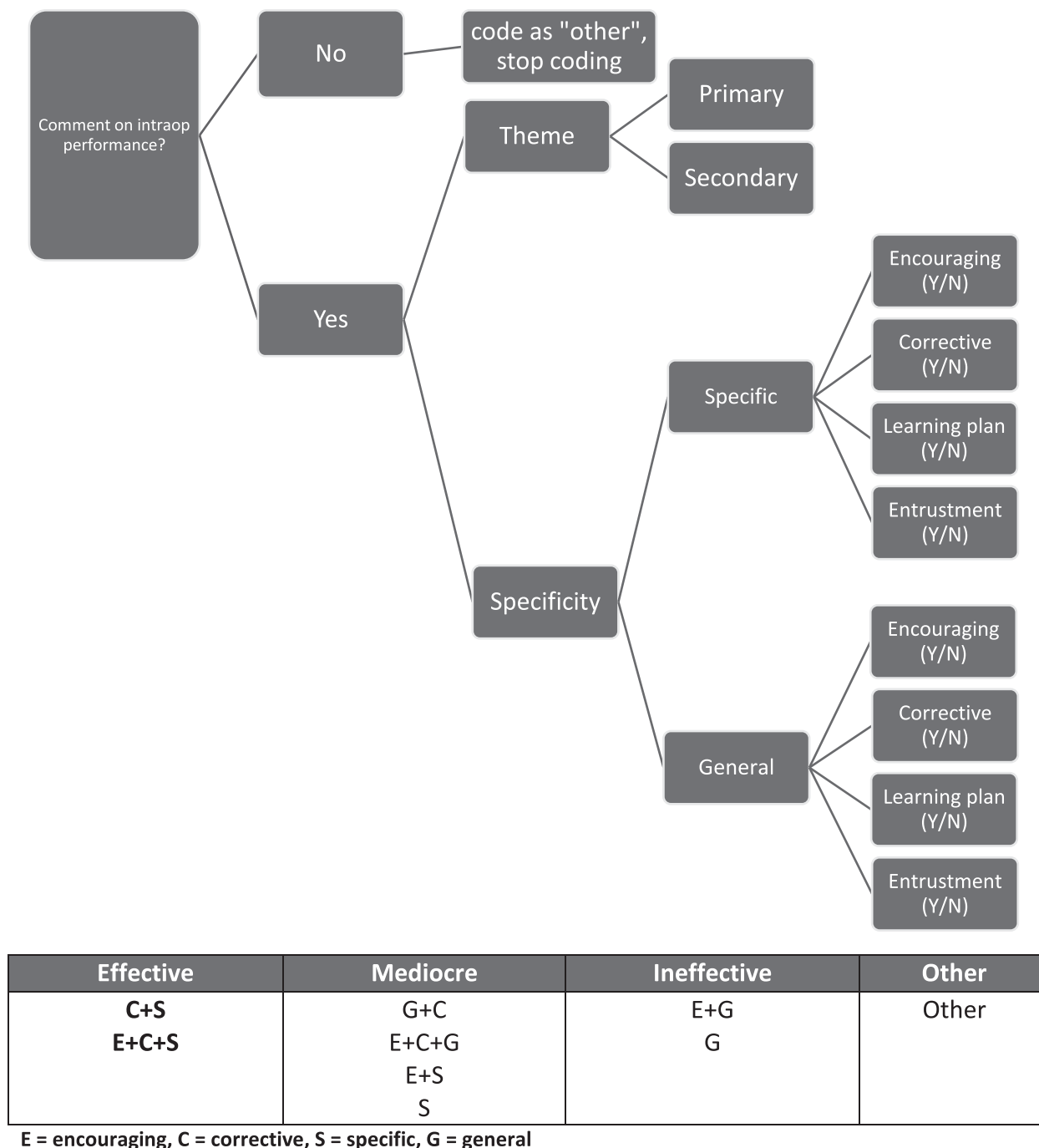
Effective	Mediocre	Ineffective	Other
C+S	G+C	E+G	Other
E+C+S	E+C+G	G	
	E+S		
	S		

E = encouraging, C = corrective, S = specific, G = general

Fig. 1. Coding framework.

trainees while providing appropriate supervision. Entrustability shifts the focus to trainees, who must demonstrate behaviors and decision making that warrant entrustment.[20] The categorization by the two surgeon raters was conducted as a deductive content analysis using a systematic approach in order to reach an inter-rater reliability of 60%, a number targeted in the literature as appropriate during qualitative analyses.[21] The transcripts were split into several data groups of roughly 50 comments each. The two surgeons rated the first set of comments independently. After the first set of comments were coded, the reviewers assigned categories and identified areas of variation. The raters explored the reason for variation and further clarified the coding scheme to improve inter-rater reliability. The same process was used after each set of data.

When needed, the raters were able to share examples of categorization inconsistencies with the rest of research team for further discussion. After all comments were categorized, the two raters reviewed each discordant categorization and agreed upon a final consensus category for analysis.

Based on the presence or absence of feedback features, narratives were graded as effective, mediocre or ineffective based on the framework proposed by Shaughness et al. (Fig. 1).[19] Pediatric surgical trainees were surveyed via an anonymous paper form, and asked to numerically assign value ratings (1 = least valued, 5 = most valued) to each dictated feedback feature. Summary statistics are provided as applicable. Data analysis was performed with IBM SPSS Statistics for Windows, version 26 (IBM Corp.,

Table 2
Dictated feedback themes.

Theme	Frequency (%) Primary Theme	Frequency (%) Secondary Theme	Definitions
Technical skill	37%	11%	A comment specifically about tissue or instrument handling, knot tying, or other technical skill
Operative Planning/Flow	15%	12%	Comments related to operative flow/anticipating next steps/transition between steps of the procedure Taking patient specific factors and imaging features to plan a procedure
Recap or Case Summary	13%	6%	Running commentary of what happened in the OR
Decision Making/Troubleshooting/ Operative Judgment	8%	8%	Specific comment on ability to resolve a difficult aspect of case Ability to resolve unexpected Ability to choose between different surgical approaches
Clinical Pearl	6%	12%	Specific tip or trick
Teaching	5%	4%	Comment on teaching skills of trainee or descriptions of how to teach a particular aspect of a procedure
Case difficulty	4%	14%	Specific comment on what made the case harder or easier
Knowledge	4%	4%	Understanding of anatomy, indication of procedure, basic steps of procedure
Other	3%	3%	
Interprofessional Collaboration/ Communication	1%	5%	Comment on communication with the surgical team or teamwork

Armonk, N.Y., USA).

Results

From June 2017 to December 2018, a total of 781 SIMPL assessments out of 1951 eligible cases (40% assessment rate) were performed by 21 faculty and 5 pediatric surgical trainees; of these, 451 had dictated comments (57% dictation rate) from 19 different faculty raters. Sixteen (4%) comments were not about intra-operative performance and were excluded. The most common primary themes were technical skill (37%), operative planning (15%) and case summary (13%); while the most common secondary theme was case difficulty (14%, Table 2). The majority of comments were encouraging (93%) and specific (65%). Only 21% of comments were corrective, with 8% of comments having an explicit learning plan. Comments related to a trainee's entrustment level were seen in 17% of instances. The median number of feedback features per comment was 2 (interquartile range 1–3). Feedback quality was considered mediocre in almost half of the comments (45%), ineffective in a third (33%) and effective only in 21%. The concordance on initial coding between the two raters was 75%.

When surveyed, trainees (n = 5) valued the most feedback that contained a learning plan (mean 4.4; range 1 [least] to 5 [most valued]), corrective features (4) or an entrustment statement (3.6), while feedback with specific (2) or encouraging (1) features was valued the least (Fig. 2).

Discussion

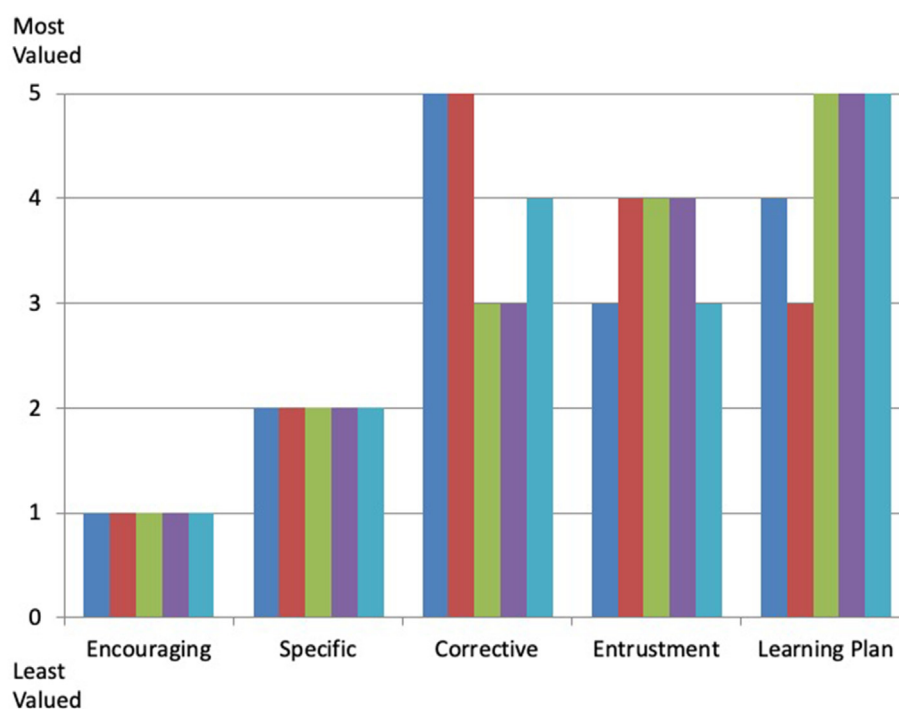Our study results suggest that slightly more than half of SIMPL

Fig. 2. Trainee's Value Ratings of Feedback Features**Each color bar denotes a different trainee (n = 5). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

evaluations of pediatric surgery trainees included dictated feedback, which tended to be encouraging and specific. Other features such as corrective feedback, entrustment elements, or learning plans were less frequently present. In fact, these later features were the most valued by our trainees, suggesting that the effectiveness of dictated feedback could be greatly improved.

Feedback has been proven to be most effective if directed at unsatisfactory elements of performance (corrective) and linked to specific learning goals (learning plan).[22] Information about "what went wrong" that fails to enable learner action toward "how you can improve" is merely "dangling data" that is unlikely to motivate learning.[23] In other words, feedback should be actionable, enabling the trainee to construct strategies for improvement. Scaffolding is a theoretical concept that works well for this purpose.[24] Scaffolding is based on supporting the learner in their zone of proximal development (beyond their current ability).[25] For trainees with a low level of competence, scaffolding involves more directive or specific feedback; while for advanced learners scaffolding can be less directive (i.e. suggestions, hints and tips for further improvement).[23] For example, a feedback comment such as: "I sense you are not quite sure how to judge the distal extent of the pyloromyotomy … next time let's focus on visualizing and palpating the duodenal shoulder or border to guide the extent of our pyloromyotomy". This feedback statement is specific, contains a learning plan, and enables a scaffolding teaching strategy by identifiying and articulating the learner's zone of proximal development and providing feedback that is learner level and context focused.

That feedback should be specific seems self-evident; nonetheless, lack of specificity has repeatedly been identified as a weakness of feedback in medical training.[26] Despite this, it is clear that general information unrelated to the performance, vague comments about good or poor performance or general compliments have been shown to be less effective than specific comments.[27,28] Encouraging feedback in and of itself is not necessarily detrimental. In fact, the theoretical model of the feedback sandwich, in which the supervisor describes what went well, what can be improved, then re-emphasizes what went well,[29] harnesses the psychological effect of praise to enable the reception of criticism. This approach is thought to be helpful especially in the delicate start of a feedback relationship, but likely unnecessary once the relationship is robust.[23]

SIMPL operative assessments are heavily framed around assessment of trainee operative autonomy; hence the incorporation of entrustment elements into feedback dictations associated with SIMPL ratings is a natural alignment. For example, based on the autonomy rating provided, faculty should be able to articulate which trainee behaviours or actions (entrustability) would lead to increased entrustment and operative autonomy during a subsequent encounter. In other words, being able to make the connection between the summative assessment and the associated formative feedback in order to generate a learning plan should be the goal of the dictated feedback.

In its current form, SIMPL assessments encourage unstructured or "ad lib" dictated feedback. Our study shows that such unstructured feedback tends to miss several key elements that could improve the quality and effectiveness of feedback. Fostering structured feedback could in turn facilitate the inclusion of such elements. Nonetheless, studies that have examined structured versus unstructured feedback have yielded mixed results. On one hand, structured debriefing after scenario-based simulation training appears to be more effective than unstructured debriefing.[30–32] On the other hand, some learners have voiced preference and attached more value to informal verbal feedback over formal work-place based assessments with written feedback.[33,34]

Limitations

Our study has limitations that we acknowledge. It evaluates data from a single training institution, and how our results apply to other learner populations and settings is not known. Our study population is composed of advanced learners and it is unlikely that our results would apply to novice learners who may receive different types of feedback features. For example, the study by Shaughness et al.[19] evaluated written feedback quality of medical students during their surgery clerkship, and found that most feedback was non-specific and encouraging, with the majority of comments being rated as ineffective. Furthermore, not all SIMPL evaluations received dictated feedback, and this "non-dictation" rate could bias our results, as faculty who consistently dictate feedback could yield different feedback features than those who provided fewer or no dictated feedback comments. Also, the dictation transcripts were de-identified and hence blinded to faculty rater and trainee, for which we were not able to evaluate feedback quality between or within faculty raters.

Implications and future directions

We envision several potential strategies that could improve the quality of dictated feedback associated with SIMPL evaluations. First, the SIMPL rater training,[18] which as of now focuses mostly on calibrating raters on the scales of operative autonomy and performance, provides no formal instruction on effective dictated feedback features, and though feedback is strongly encouraged, it remains optional. Hence, incorporating training on effective feedback features during SIMPL rater training should be considered. Second, the SIMPL app itself could be modified to prompt raters to include effective feedback elements in their dictation. Third, this type of prompting or reminders could be also incorporated in the daily email reminders which faculty raters receive for pending evaluations. Furthermore, faculty development sessions should focus on highlighting which feedback features are more effective, provide examples of feedback dictations of varied quality, and point out which feedback elements are most valued by trainees. Though the focus on feedback has largely been on how faculty should deliver feedback, future research is needed to understand how learners receive and respond to feedback, and how institutions can create a culture in which feedback works optimally.[23]

Conclusions

Dictated feedback associated with SIMPL operative assessments of pediatric surgical trainees is mostly of mediocre quality and contains predominantly encouraging and specific feedback features. Trainees value feedback that is corrective, contains entrustment features and learning plans over feedback that is encouraging or specific. In order to achieve greater quality and potentially much more effective dictated feedback, faculty development efforts should focus on aligning these feedback features and trainee preferences, using strategies aimed at enhancing the delivery of corrective feedback, incorporating entrustment elements and generating learning plans.

References

1. Taylor DCM, Hamdy H. *Adult Learning Theories: Implications for Learning and Teaching in Medical Education: AMEE Guide No. 83*. Med Teach; 2013. https://doi.org/10.3109/0142159X.2013.828153.
2. Van De Ridder JMM, Stokking KM, McGaghie WC, Ten Cate Otj. What is feedback in clinical education? *Med Educ*. 2008. https://doi.org/10.1111/j.1365-2923.2007.02973.x.
3. Hattie J, Timperley H. The power of feedback. *Rev Educ Res*. 2007. https://

doi.org/10.3102/003465430298487.

4. Hatala R, Cook DA, Zendejas B, Hamstra SJ, Brydges R. Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. *Adv Health Sci Educ Theory Pract*. 2014;19(2):251–272. https://doi.org/10.1007/s10459-013-9462-8.

5. Mulliner E, Tucker M. Feedback on feedback practice: perceptions of students and academics. *Assess Eval High Educ*. 2017. https://doi.org/10.1080/02602938.2015.1103365.

6. Karim AS, Sternbach JM, Bender EM, Zwischenberger JB, Meyerson SL. Quality of operative performance feedback given to thoracic surgery residents using an app-based system. *J Surg Educ*. 2017;74(6):e81–e87. https://doi.org/10.1016/j.jsurg.2017.10.001.

7. Eaton MJ, Schuller MC, Yang A, et al. Value and barriers to use of the SIMPL tool for resident feedback. J Surg Educ. 76(3):620–627.

8. Ten Cate O, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad Med.*. October 2018:1 https://doi.org/10.1097/ACM.0000000000002495.

9. Cogbill TH, Ashley SW, Borman KR, Buyske J, Cofer JB, Deladisma AM. *The General Surgery Milestone Project*; 2014. https://www.acgme.org/acgmeweb/Portals/0/PDFs/Milestones/SurgeryMilestones.pdf.

10. Norcini J, Zaidi Z. *Workplace assessment: evidence, theory, and practice*. 2018: 319–334 https://doi.org/10.1002/9781119373780.ch22.

11. Norcini J, Burch V. *Workplace-based Assessment as an Educational Tool: AMEE Guide No. 31*. Med Teach; 2007. https://doi.org/10.1080/01421590701775453.

12. George BC, Bohnen JD, Williams RG, et al. Readiness of US general surgery residents for independent practice. *Ann Surg*. 2017;266(4):582–594. https://doi.org/10.1097/SLA.0000000000002414.

13. Darosa DA, Zwischenberger JB, Meyerson SL, et al. A theory-based model for teaching and assessing residents in the operating room. *J Surg Educ*. 2013;70(1):24–30. https://doi.org/10.1016/j.jsurg.2012.07.007.

14. Kobraei EM, Bohnen JD, George BC, et al. Uniting evidence-based evaluation with the ACGME plastic surgery milestones. *Plast Reconstr Surg*. 2016;138(2): 349e–357e. https://doi.org/10.1097/PRS.0000000000002411.

15. Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intraoperative performance assessment with SIMPL (system for improving and measuring procedural learning): early experience from a multi-institutional trial. *J Surg Educ*. 2016;73(6):e118–e130. https://doi.org/10.1016/j.jsurg.2016.08.010.

16. George BC, Teitelbaum EN, Meyerson SL, et al. Reliability, validity, and feasibility of the zwisch scale for the assessment of intraoperative performance. *J Surg Educ*. 2014;71(6):e90–e96. https://doi.org/10.1016/j.jsurg.2014.06.018.

17. Zendejas B, Lillehei CW, George BC, Modi BP. Assessment of operative autonomy and readiness for independent practice among pediatric surgery fellows. *J Pediatr Surg*. 2020. https://doi.org/10.1016/j.jpedsurg.2019.09.061.

18. George BC, Teitelbaum EN, DaRosa DA, et al. Duration of faculty training needed to ensure reliable or performance ratings. *J Surg Educ*. 2013;70(6): 703–708. https://doi.org/10.1016/j.jsurg.2013.06.015.

19. Shaughness G, Georgoff PE, Sandhu G, et al. Assessment of clinical feedback given to medical students via an electronic feedback system. *J Surg Res*. 2017. https://doi.org/10.1016/j.jss.2017.05.055.

20. Sandhu G, Thompson-Burdine J, Nikolian VC, et al. Association of faculty entrustment with resident autonomy in the operating room. *JAMA Surg*. 2018. https://doi.org/10.1001/jamasurg.2017.6117.

21. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. *Int J Qual Methods*. 2020. https://doi.org/10.1177/1609406919899220.

22. van de Ridder JMM, Mcgaghie WC, Stokking KM, ten Cate Otj. Variables that affect the process and outcome of feedback, relevant for medical training: a meta-review. *Med Educ*. 2015. https://doi.org/10.1111/medu.12744.

23. Lefroy J, Watling C, Teunissen PW, Brand P. Guidelines: the do's, don'ts and don't knows of feedback for clinical education. *Perspect Med Educ*. 2015. https://doi.org/10.1007/s40037-015-0231-7.

24. Wood D, Bruner JS, Ross G. The role OF tutoring IN problem solving. *JCPP (J Child Psychol Psychiatry)*. 1976. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x.

25. Vygotsky L. Thought and language - revised edition. *J Appl Behav Sci*. 1986: 12–58.

26. Weston PSJ, Smith CA. The use of mini-CEX in UK foundation training six years following its introduction: lessons still to be learned and the benefit of formal teaching regarding its utility. *Med Teach*. 2014. https://doi.org/10.3109/0142159X.2013.836267.

27. Shute VJ. Focus on formative feedback. *Rev Educ Res*. 2008. https://doi.org/10.3102/0034654307313795.

28. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull*. 1996. https://doi.org/10.1037/0033-2909.119.2.254.

29. Dohrenwend A. Serving up the feedback sandwich. *Fam Pract Manag*. 2002;9(10):43–46.

30. Arora S, Ahmed M, Paige J, et al. Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. *Ann Surg*. 2012. https://doi.org/10.1097/SLA.0b013e3182610c91.

31. Nyström S, Dahlberg J, Edelbring S, Hult H, Abrandt Dahlgren M. Debriefing practices in interprofessional simulation with students: a sociomaterial perspective. *BMC Med Educ*. 2016. https://doi.org/10.1186/s12909-016-0666-5.

32. Cheng A, Eppich W, Grant V, Sherbino J, Zendejas B, Cook DA. Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. *Med Educ*. 2014;48(7):657–666. https://doi.org/10.1111/medu.12432.

33. Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. *Med Educ*. 2013. https://doi.org/10.1111/medu.12087.

34. Urquhart LM, Rees CE, Ker JS. Making sense of feedback experiences: a multi-school study of medical students' narratives. *Med Educ*. 2014. https://doi.org/10.1111/medu.12304.