

# Feedback-Seeking Behavior and Practice Readiness for General Surgery



John Luckoski, MD,\* Angela Thelen, MD,\* Dylan Russell, MD,<sup>†</sup> Brian George, MD, MAEd,\* and Andrew Krumm, PhD<sup>‡</sup>, on behalf of the Society for Improving Medical Professional Learning (SIMPL)

\*Department of Surgery, University of Michigan, Ann Arbor, Michigan; <sup>†</sup>Department of Surgery, Tripler Army Medical Center, Honolulu, Hawaii; and <sup>‡</sup>Department of Learning Health Sciences, University of Michigan, Ann Arbor, Michigan

**OBJECTIVE:** To analyze the relationship between feedback-seeking behavior, operationalized as the number of trainee-requested evaluations, with ratings of surgical trainees' operative autonomy and performance.

**DESIGN:** We analyzed operative assessment data using the System for Improving and Measuring Procedural Learning's smartphone-based assessment app called Society for Improving Medical Professional Learning (SIMPL) OR. Using cross-classified mixed effects models, we analyzed the association between trainee-requested SIMPL OR app evaluations and both trainee performance and autonomy ratings. Models included covariates for requested evaluations, PGY-year, month of the academic year, and patient-related case complexity. Random effects for program, procedure, rater, and trainee were also included to account for correlations among evaluations. Only ratings for procedures deemed Core to general surgery were included.

**SETTING:** Operative assessment data using the SIMPL OR app requested by categorical U.S. general surgery residents between September 2015 to April 2021.

**PARTICIPANTS:** A total of 61 general surgery residency programs, encompassing 2190 categorical general surgery residents.

**RESULTS:** A total of 58,104 SIMPL app operative assessments were analyzed. Autonomy scores were weakly but positively associated with number of trainee-requested evaluations ( $B = 0.002$ ,  $p < 0.001$ ). Trainee-requested evaluations were also statistically associated with operative performance scores ( $B = 0.002$ ,  $p < 0.001$ ).

**Funding source:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Correspondence:** Inquiries to John Luckoski, MD, Center for Surgical Training and Research, University of Michigan, 2800 Plymouth Road (North Campus Research Complex), Building #10 Room A193, Ann Arbor, MI, 48109; e-mail: Luckoski@med.umich.edu

**CONCLUSIONS:** The propensity of a resident to seek feedback using the SIMPL app was weakly associated with higher operative autonomy ratings and higher operative performance ratings. While regular feedback is important for monitoring performance over time, more direct approaches related to the quality of feedback that trainees receive may be needed to better assess the relationships between feedback-seeking behavior and operative autonomy as well as performance. (J Surg Ed 79:295–301. © 2021 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

**KEY WORDS:** surgical education, assessment, autonomy, workplace-based assessment, general surgery residency, resident performance

**COMPETENCIES:** Practice-Based Learning and Improvement

## INTRODUCTION

While workplace-based assessments (WBAs) are widely seen as essential building blocks of graduate medical education assessment systems,<sup>1</sup> effective implementation remains challenging.<sup>2–7</sup> In WBA systems that largely rely on trainees to initiate an assessment, less engaged trainees, by definition, generate fewer evaluations. Fewer evaluations lead to fewer opportunities for formal, documented feedback. In this paper, we characterize trainees requesting an evaluation as feedback-seeking behavior. Research in nonmedical domains has demonstrated a positive association between feedback-seeking and performance.<sup>8–10</sup> Yet, it is unclear how feedback-seeking is associated with a trainee's performance in general surgery.

A popular WBA system within general surgery is the smartphone-based assessment app from the Society for Improving Medical Professional Learning (SIMPL).<sup>11</sup>

Trainees who desire feedback can use the SIMPL app to request performance ratings and narrative comments from their supervising faculty. Given the trainee-driven approach for how the SIMPL app is implemented in many programs, it is an ideal testbed for examining the potential relationship between trainees' feedback-seeking and measures of operative autonomy and performance.

Understanding the role of feedback-seeking behavior could help program directors better understand what contributes to ratings of trainees' operative autonomy and performance over and above well-documented factors such as trainees' postgraduate year (PGY), differences among raters, differences between procedures, and the overall complexity of the case. Moreover, a better understanding of the effect of feedback-seeking could motivate future study of the quality, and not just quantity, of feedback being delivered to trainees. To better understand the role of feedback-seeking behavior, we analyzed the relationship between the number of trainee-requested evaluations and ratings of operative autonomy and performance, controlling for factors known to be strongly and significantly related to the ratings that trainees receive.

## METHODS AND MATERIALS

### Data Sources

The SIMPL app assessment data was obtained directly from the Society for Improving Medical Professional Learning, a nonprofit consortium which maintains the database of SIMPL app evaluations for research and quality improvement purposes. We collected SIMPL app assessments from September 2015 through April 2021. The SIMPL app allows attending surgeons and trainees to independently rate a trainee's performance on a given operative case up to 72 hours after case completion. The attending surgeon and trainee are asked to rate the trainee's autonomy using the Zwisch scale, operative performance using a performance scale, and relative case complexity. Raters can also provide narrative comments for each evaluation.

The Zwisch scale assesses autonomy on a 4-level rating scale ranging from "Show and Tell" to "Supervision Only." Each rating-level increase equates to a trainee requiring less direction and being granted more autonomy. The performance scale prompts faculty to rate a trainee's operative skill during the case on a 5-point scale from "Unprepared/Critical Deficiency" to "Exceptional Performance." A trainee who is given a 1 for autonomy does not receive a performance rating, therefore the number of evaluations differ between models. The complexity scale asks raters to evaluate if the case was

relatively more or less complex than a typical case of the same type. Details of these rating scales and evidence regarding their validity and reliability have been previously published in detail.<sup>11,12</sup>

### Study Population and Procedural Taxonomy

Our study population included all categorical U.S. general surgery residents who submitted a SIMPL app evaluation for a General surgery procedure. General surgery procedures were defined as all procedures deemed "Core" or "Advanced" procedures based on the American Board of Surgery's 2020 Surgical Council on Resident Education (SCORE) curriculum guidelines.<sup>12</sup> Procedures that were not labeled Core or Advanced within the SCORE curriculum, such as gynecologic procedures and genitourinary procedures within the SCORE taxonomy were excluded. Additionally, given the focus of this study on trainee feedback-seeking behaviors, only assessments in which the resident initiated the evaluation process were included in the analysis of a trainee's number of assessments, but their overall performance and autonomy ratings utilized all assessments completed.

### Statistical Analysis

Using a cross-classified linear mixed effects model, we analyzed the association between the volume of resident-initiated SIMPL app assessments (i.e., Requests) and trainee autonomy and operative performance. Requests was modeled as a trainee-level fixed effect, which means that a trainee's total number of requested evaluations was repeated for each individual SIMPL app evaluation. The way in which we operationalized requests models trainees' total number of requests from the outset, which captures the idea that some trainee's evidence feedback seeking behavior during both the first and last evaluation. This operationalization differs from using a growth term that changes with each request.

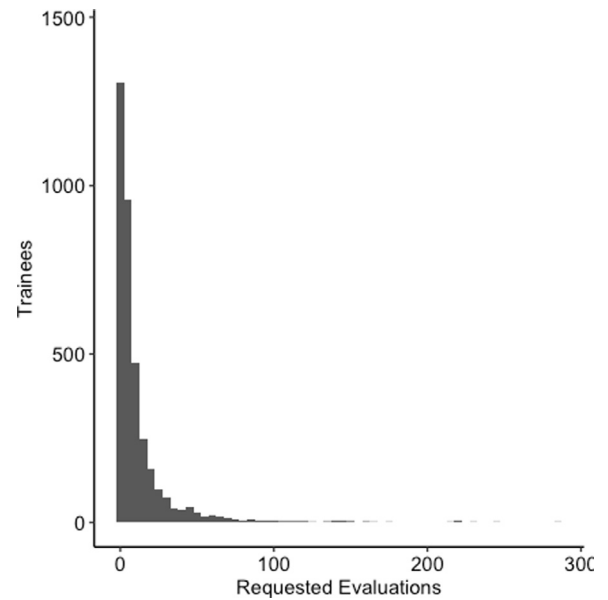
Random effects included trainee, rater, procedure, and program. Models included covariates for the gender of both raters and trainees, PGY, month of the academic year, interaction between PGY and requests, and patient-related case complexity. Gender was included to control for potential bias of attending surgeons; male was treated as the reference category. Month of academic year was included to control for changes in average operative autonomy and performance that occur over time within a year, and PGY was included to account for changes that occur across years. Month was modeled as a continuous variable whereby July was set to 0 and June was set to 11. PGY was modeled as a categorical variable, and PGY 1 was used as the reference category. We included interactions between PGY and

requests to account for average differences in requests across PGYs. Lastly, a 3-level case complexity variable was included with “Average Complexity” used as the reference category. Separate models were run for operative autonomy and performance ratings, using the numeric version of each variable. For our primary results, requests were modeled as a cumulative count per trainee; for an additional set of analyses, we group-mean centered cumulative requests within a program to account for average differences between programs because some programs require trainees to log more procedures whereas other programs let the trainee drive the feedback requests process almost entirely (See Supplementary Material). All statistical analyses were performed using R version 4.0.2 with the lme4 package.<sup>13</sup> This study was approved by the Institutional Review Board at the University of Michigan.

## RESULTS

Across 61 general surgery residency programs, 2190 categorical general surgery residents requested 58,104 SIMPL app operative assessments for general surgery procedures over a 5-year period. Seventy-nine percent of evaluations were requested by trainees, and approximately 21% were initiated by faculty. These operative assessments spanned 947 SIMPL app procedure codes. **Table 1** provides a descriptive overview of trainees and evaluations by PGY levels. **Figure 1** demonstrates the distribution of requested evaluations across all PGYs with an overall mean of 11.6 (SD 21.4) and a median of 5.

For the primary models where the autonomy and performance scales were used as dependent variables, there were significant, positive associations between increased operative autonomy ratings and the number of Requests made by trainees ( $B = 0.002$ ) as well as for performance ratings ( $B = 0.002$ ,  $p < .001$ ). Summary of these findings are displayed in **Table 2**. Across both models, strong and significant predictors included the overall complexity of the case, whereby High Complexity cases



**FIGURE 1.** Histogram of requested evaluations by trainee (Note: 0 requested evaluations indicates trainees who only had evaluations initiated by an attending surgeon; bin width = 5).

were, on average, 0.29 Autonomy scale points lower and 0.101 Performance scale points lower, respectively, than Average Complexity cases. We also observed a negative effect for the female trainees in relation to the average autonomy they are granted ( $B = -0.027$ ,  $p < .001$ ).

PGY level provided some of the strongest effects for both Autonomy and Performance, and changes in months within a PGY year, on average, were significant across both models as well. The effects of Trainee, Rater, and Procedure were also strong. For example, a trainee who was 1 standard deviation above average is likely to score 0.18 Autonomy scale points higher and 0.19 Performance scale points higher. Likewise, a trainee who was scored by a rater who was 1 standard deviation above average is likely to be scored 0.31 Autonomy scale points higher. The group-means centered model produced similar results, indicating that trainees who are above average in their respective in terms of the number

**TABLE 1.** Descriptive Information of Assessment Cases Stratified by PGY Status

	PGY 1	PGY 2	PGY 3	PGY 4	PGY 5
Mean/Median cases per Trainee (SD)	7.80/4 (11.4)	9.56/5 (14.3)	12.7/5 (24.2)	13.50/5 (24.4)	13.70/5 (25.3)
Mean Autonomy scale (SD)	1.99 (0.713)	2.34 (0.752)	2.57 (0.786)	2.82 (0.801)	3.05 (0.786)
Mean Performance scale (SD)	2.75 (0.656)	3.05 (0.653)	3.28 (0.658)	3.57 (0.690)	3.86 (0.611)
Mean Complexity scale (SD)	1.97 (0.563)	2.05 (0.577)	2.08 (0.592)	2.14 (0.605)	2.18 (0.609)

Subjects:  $n = 2190$  Cases:  $n = 58,104$ .

**TABLE 2.** Fixed and Random Effects on Autonomy and Performance Scores

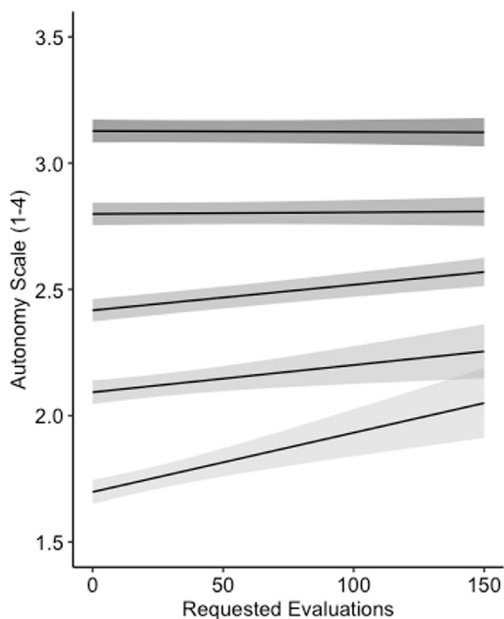
Fixed Effects	Autonomy			Performance		
	B	Std. Error		B	Std. Error	
Intercept	1.615	0.0262		2.487	0.0245	
Requests	0.002	0.0005		0.002	0.0005	
Month	0.018	0.0009		0.019	0.0008	
Trainee Gender	-0.027	0.0114		-0.007*	0.0111	
Rater Gender	-0.012*	0.0204		0.022*	0.0212	
Low Complexity	0.119	0.0082		0.104	0.0072	
High Complexity	-0.287	0.0065		-0.101	0.0058	
PGY 2	0.395	0.0169		0.333	0.0160	
PGY 3	0.719	0.0162		0.625	0.0154	
PGY 4	1.101	0.0169		1.027	0.0160	
PGY 5	1.430	0.0173		1.371	0.0164	
PGY 2 X Requests	-0.001	0.0006		-0.001*	0.0005	
PGY 3 X Requests	-0.001	0.0005		-0.001*	0.0005	
PGY 4 X Requests	-0.002	0.0005		-0.002	0.0005	
PGY 5 X Requests	-0.002	0.0005		-0.002	0.0005	
Random Effects	Var.	SD	ICC	Var.	SD	ICC
Residual	0.354	0.59	53%	0.252	0.50	52%
Trainee	0.033	0.18	5%	0.035	0.19	7%
Rater	0.098	0.31	14%	0.114	0.34	24%
Procedure	0.185	0.43	27%	0.077	0.28	16%
Program	0.003	0.05	< 1%	0.006	0.08	1%

\*p values >0.05; all other p values significant <0.05./ICC, Intraclass correlation.

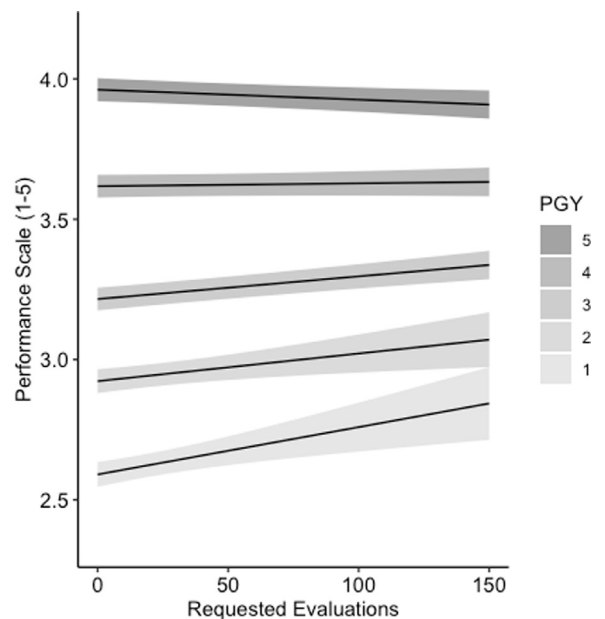
of feedback requests also score more highly on the Autonomy and Performance scales, respectively.

Figures 2 and 3 present the adjusted predictions, or marginal effects, for trainee requested evaluations by PGY on the Autonomy and Performance scales,

respectively. These figures more clearly illustrate the effects of the PGY X Requests interaction and the ways in which lower PGY levels experience more benefit from increased feedback seeking for both Autonomy and Performance.



**FIGURE 2.** Autonomy rating by requested evaluations and PGY.



**FIGURE 3.** Performance rating by requested evaluations and PGY.

## DISCUSSION

Trainees who exhibited more feedback-seeking behavior were more likely to receive increased autonomy ratings across all Core surgical cases. Similarly, there was an association between feedback-seeking behavior and operative performance ratings in the SIMPL app. Our results reiterated previously known findings that both autonomy and operative performance ratings are strongly influenced by PGY level and case complexity as well as the trainee, rater, and procedure.<sup>12</sup>

The association between feedback-seeking behavior and autonomy is supported within the surgical education literature. Woelfel et al. defined a 3-stage process for trainees to gain operative autonomy that begins with building rapport. They specifically described 4 common strategies for advancing operative autonomy including “smart communication, attention to attending preferences, helpful allies, and visible attributes.”<sup>14</sup> Similarly, requesting SIMPL app evaluations may function as a smart communication strategy to build rapport between a trainee and their supervising faculty, allowing the trainee to gain higher levels of operative autonomy. This likely contributes to our findings that trainees who more frequently request feedback are more likely to achieve slightly higher levels of autonomy.

The association between feedback-seeking behavior and performance ratings is also consistent with prior work in this domain.<sup>15-18</sup> Veloski et al. reported that the main effects of formal feedback on performance are influenced in part by the duration of feedback, as well as the physicians’ active involvement in the process. Gaunt et al. evaluated feedback-seeking behavior in general surgery resident and found that trainees’ feedback-seeking behavior is in part related to the trainees’ perceived value of the assessments as well as the supportiveness of their supervising faculty.<sup>19,20</sup> In other words, if a trainee’s perception of a WBA is that it does not represent an opportunity to learn and improve, then the potential benefits of feedback may not be experienced by a trainee. Overall, the association between feedback-seeking and operative autonomy as well the association between feedback-seeking and performance necessitates a deeper exploration for why and how trainees are seeking feedback and the quality of the feedback provided by supervising surgeons.

### Limitations

There are several limitations to this study. First, we only utilized SIMPL app data, which may limit generalizability to other WBAs. Second, the typical number of SIMPL app requests was relatively low (Median = 5) and highly variable across trainees, although our analyses included trainee random effects to account for clustering. Third, a significant amount of variation in both autonomy and

performance ratings can be attributed to the PGY level, case complexity, rater, trainee, and procedure, making it difficult to identify the unique contribution of feedback-seeking behavior. Finally, we must be cautious when using the number of trainee-requested evaluations as a proxy for feedback-seeking behavior. Simply requesting feedback, alone, is not likely to contribute to trainees’ development, the quality of the feedback and the degree to which trainees are influenced by that feedback are the primary mechanisms by which feedback is thought to support increased performance over time.<sup>6</sup>

### Impact

The associations between trainee-requested feedback and operative autonomy and performance warrants further scrutiny. While multiple factors are clear drivers of SIMPL app ratings (e.g., PGY and rater), the relationship between requested evaluations and increased performance and autonomy is strong evidence for further exploring how feedback can be made as beneficial as possible.

### Next Steps

While increased feedback-seeking behavior with the SIMPL app correlates with enhanced autonomy and performance, the mechanisms driving this effect remain unclear. An important next step for this area of inquiry is triangulating feedback seeking in the SIMPL app with other data sources related to the ways in which trainees seek feedback. An ideal feedback system is one that actively helps a trainee improve and there are myriad ways for trainees to receive such feedback. Within the context of the SIMPL app, a further avenue to explore is this relationship between the quality of feedback provided by the SIMPL app in the form of narrative comments. Gathering data from multiple sources on the quantity and quality of feedback will provide a deeper understanding of feedback seeking in the training of surgeons.

## CONCLUSIONS

The propensity of a resident to seek feedback using the SIMPL app was weakly associated with higher operative autonomy ratings and higher operative performance ratings. While regular feedback is important for monitoring performance over time, more direct approaches related to the quality of feedback that trainees receive may be needed to better assess the relationships between feedback-seeking behavior and operative autonomy as well as performance.

## CONFLICT OF INTEREST

None.

## REFERENCES

1. Prescott-Clements L, van der Vleuten CP, Schuwirth LW, Hurst Y, Rennie JS. Evidence for validity within workplace assessment: the Longitudinal Evaluation of Performance (LEP). *Med Educ*. 2008;42:488–495. <https://doi.org/10.1111/j.1365-2923.2007.02965.x>.
2. Aryal KR, Currow C, Downey S, Praseedom R, Seager A. Work-based assessments in higher general surgical training program: a mixed methods study exploring trainers' and trainees' views and experiences. *Surg J*. 2020;6:e49–e61. <https://doi.org/10.1055/s-0040-1708062>.
3. James K, Cross K, Lucarotti M, Fowler A, Cook T. Undertaking procedure-based assessment is feasible in clinical practice. *Ann R Coll Surg Engl*. 2009;91:110–112. <https://doi.org/10.1308/003588409x359286>.
4. Gaunt A, Patel A, Rusius V, Royle TJ, Markham DH, Pawlikowska T. Playing the game”: how do surgical trainees seek feedback using workplace-based assessment? *Med Educ*. 2017;51:953–962. <https://doi.org/10.1111/medu.13380>.
5. Jong LH de, Favier RP, Vleuten CPM van der, Bok HGJ. Students' motivation toward feedback-seeking in the clinical workplace. *Med Teach*. 2017;39:954–958. <https://doi.org/10.1080/0142159X.2017.1324948>.
6. McKendy KM, Watanabe Y, Lee L, et al. Perioperative feedback in surgical training: a systematic review. *Am J Surg*. 2017;214:117–126. <https://doi.org/10.1016/j.amjsurg.2016.12.014>.
7. 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:e118–e130. <https://doi.org/10.1016/j.jsurg.2016.08.010>.
8. Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ*. 2012;17:15–26. <https://doi.org/10.1007/s10459-011-9290-7>.
9. Crommelinck M, Anseel F. Understanding and encouraging feedback-seeking behaviour: a literature review. *Med Educ*. 2013;47:232–241. <https://doi.org/10.1111/medu.12075>.
10. Ridder JMM van de, McGaghie WC, Stokking KM, Cate OTJ ten. Variables that affect the process and outcome of feedback, relevant for medical training: a meta-review. *Med Educ*. 2015;49:658–673. <https://doi.org/10.1111/medu.12744>.
11. George BC, Bohnen JD, Schuller MC, Fryer JP. Using smartphones for workplace-based performance assessment: a SIMPL case study. *Surgery*. 2020;167:903–906.
12. George BC, Bohnen JD, Williams RG, et al. Readiness of US general surgery residents for independent practice. *Ann Surg*. 2017;266:582–594.
13. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
14. Woelfel I, Smith BQ, Strosberg D, et al. Residents' method for gaining operative autonomy. *Am J Surg*. 2020;220:893–898. <https://doi.org/10.1016/j.amjsurg.2020.03.022>.
15. De Stobbeleir KEM, Ashford S, Sully de Luque M. How is Feedback-Seeking Behavior Interpreted? The Influence of Feedback-Seeking Pattern and Feedback Source as Characteristics on Impression Formation and Performance Evaluations. Ghent Univ Fac Econ Bus Adm Work Pap Fac Econ Bus Adm Ghent Univ Belg; 2008 Published online.
16. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach*. 2006;28:117–128. <https://doi.org/10.1080/01421590600622665>.
17. Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Adv Health Sci Educ*. 2016;21:455–473. <https://doi.org/10.1007/s10459-015-9614-0>.
18. Phillips AW, Madhavan A, Bookless LR, Macafee DAL. Surgical trainers' experience and perspectives on workplace-based assessments. *J Surg Educ*. 2015;72:979–984. <https://doi.org/10.1016/j.jsurg.2015.03.015>.
19. Gaunt A, Patel A, Fallis S, et al. Surgical trainee feedback-seeking behavior in the context of workplace-

based assessment in clinical settings. *Acad Med.* 2017;92:827-834. <https://doi.org/10.1097/ACM.0000000000001523>.

20. Gaunt A, Markham DH, Pawlikowska TRB. Exploring the role of self-motives in postgraduate trainees'

feedback-seeking behavior in the clinical workplace: a multicenter study of workplace-based assessments from the United Kingdom. *Acad Med.* 2018;93:1576. <https://doi.org/10.1097/ACM.0000000000002348>.

## SUPPLEMENTARY INFORMATION

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.jsurg.2021.10.003](https://doi.org/10.1016/j.jsurg.2021.10.003).