Evaluating the Quality of Written Feedback Within Entrustable Professional Activities in an Internal Medicine Cohort

Lorenzo Madrazo, MD
Jennifer DCruz, MD, MSc
Natasha Correa, MD
Klajdi Puka, PhD
Sheri-Lynn Kane, MD, FRCPC

ABSTRACT

Background Whether written comments in entrustable professional activities (EPAs) translate into high-quality feedback remains uncertain.

Objective We aimed to evaluate the quality of EPA feedback completed by faculty and senior residents.

Methods Using retrospective descriptive analysis, we assessed the quality of feedback from all EPAs for 34 first-year internal medicine residents from July 2019 to May 2020 at Western University in London, Ontario, Canada. We assessed feedback quality on 4 domains: timeliness, task orientation, actionability, and polarity. Four independent reviewers were blinded to names of evaluators and learners and were randomized to assess each EPA for the 4 domains. Statistical analyses were completed using R 3.6.3. Chi-square or Fisher's exact test and Cochran-Armitage test for trend were used to compare the quality of feedback provided by faculty versus student assessors, and to compare the effect of timely versus not timely feedback on task orientation, actionability, and polarity.

Results A total of 2471 EPAs were initiated by junior residents. Eighty percent (n=1981) of these were completed, of which 61% (n=1213) were completed by senior residents. Interrater reliability was almost perfect for timeliness (κ=0.99), moderate for task orientation (κ=0.74), strong for actionability (κ=0.81), and moderate for polarity (κ=0.62). Of completed EPAs, 47% (n=926) were timely, 85% (n=1697) were task oriented, 83% (n=1649) consisted of reinforcing feedback, 4% (n=79) contained mixed feedback, and 12% (n=240) had neutral feedback. Thirty percent (n=595) were semi- or very actionable.

Conclusions The written feedback in the EPAs was task oriented but was neither timely nor actionable. The majority of EPAs were completed by senior residents rather than faculty.

Introduction

Residency training in Canada has shifted to competency-based medical education (CBME) to restructure curricula around physician competencies and better prepare clinicians to serve patients.[1] This transition introduced entrustable professional activities (EPAs), which are specialty-specific clinical tasks that can be entrusted to trainees once they demonstrate competence in completing the task independently.[2,3] Because tasks outlined by EPAs are always contextualized by an assessment, we use the term EPA to encompass both. EPAs are distinct from the traditional in-training evaluation report (ITER), which is an overall rotation-based summative evaluation.[4-6] EPAs are meant to both increase and capture formative, timely, and task-specific[7] feedback in addition to existing ITERs. One of the goals of CBME is to provide opportunities for feedback and coaching for residents. However, it is unclear whether use of EPAs results in high-quality feedback. From the literature, high-quality feedback is timely, task oriented, and actionable.[8,9] Moreover, feedback that is corrective tends to provide more useful information than those that are positive or neutral in their polarity.[10]

Few studies have assessed the quality of written feedback captured through EPAs. Two studies from Queen's University School of Medicine medical oncology program showed that both faculty and residents valued high-quality written feedback captured in EPAs.[7] Their pilot study showed that 33% of feedback from 17 EPAs analyzed were actionable,[7] which later increased in a follow-up study showing that 56% of 157 EPA feedback was actionable,[11] suggesting an increased prevalence of actionable feedback over time. In a different center, a psychiatry residency program evaluated a newly implemented mobile app to facilitate EPAs and found that 95% (94 of 99) of comments were task specific.[12] Additionally,

DOI: http://dx.doi.org/10.4300/JGME-D-22-00222.1

Editor's Note: The online version of this article contains the 5-point rating scale for entrustability and sections for narrative feedback, and an analysis of the entrustable professional activities.

focus groups composing residents from multiple specialties at McMaster University revealed a perceived higher frequency of feedback with EPAs but poor-quality feedback as a result of "assessment fatigue."[13] These studies in smaller programs suggest a mixed quality of feedback received through EPAs. Ascertaining the quality of EPA feedback from larger residency programs, where CBME implementation is likely to be most challenging, may better gauge practical application of EPAs more generally. A standardized approach to assessing feedback quality was also lacking. The Canadian Excellence in Residency Accreditation requires demonstration of ongoing continuous quality improvement (CQI) program initiatives. As part of this CQI initiative, and with the implementation of CBME for internal medicine (IM) programs across Canada in July 2019,[14] our objective was to assess the quality of the written feedback being documented within EPAs for postgraduate year (PGY)-1 IM residents at Western University in the first year of CBME implementation. Specifically, our study sought to examine EPA feedback for timeliness, task orientation, actionability, and polarity and differences between feedback provided by faculty members and senior residents. In doing so, we also sought to develop a method of feedback analysis that would be translatable to other institutions using EPAs.

Objectives
The purpose of this study was to evaluate the quality of entrustable professional activity (EPA) feedback completed by faculty and senior residents.

Findings
The written feedback in the EPAs was task oriented but was neither timely nor actionable; most were completed by senior residents rather than faculty.

Limitations
Study findings were from a single program and institution which may limit generalizability.

Bottom Line
This study offers an approach to assessing the quality of written EPA feedback that can be adapted to other institutions that implement EPAs.

Methods

Setting

The IM program at Western University in London, Ontario, Canada began a preliminary implementation of CBME in the 2018-2019 academic year. This included faculty and resident education throughout the year on the use and purpose of EPAs as well as education regarding the qualities of high-quality feedback. Faculty development also included meetings with each division and out-of-town elective sites in the lead-up year to provide specialty-specific education and examples. Official implementation commenced on July 1, 2019.

EPAs at our institution are requested by a junior resident and completed by an assessor (senior resident or attending physician) electronically through an online platform (Elentra) accessed through computers or mobile devices. Residents must be assessed on a specific number of each EPA as required by the Royal College of Physicians and Surgeons of Canada.[15] Only faculty or residents more senior to the learner may complete EPAs. Each EPA has a 5-point rating scale for entrustability and has 2 sections available for narrative feedback: (1) *Comments* where assessors can provide feedback on learner performance on the task in question, and (2) *Next Steps* where assessors can provide recommendations for further development (see online supplementary data). These sections were the focus of our analyses on feedback quality. There are 10 unique EPAs across the 2 stages of training in the first year (see online supplementary data). There are multiple contextual variables for each EPA, and residents are required to obtain multiple observations.

Study Population

We analyzed all EPAs completed between July 2019 and May 2020 for 34 PGY-1 residents in the IM program.

Feedback Analysis

From the literature, we reviewed several examples of feedback analysis[7,10,12,16] to assess the quality of written feedback in the EPAs. We identified and defined the following domains as important and measurable qualities of good feedback: timeliness, task orientation, actionability, and polarity. Modification and final agreement of variable definitions was achieved after adjudication of a test set of 30 EPAs independently graded by all 4 investigators (L.M., N.C., J.D., S.K.).

Timely feedback was defined as EPA completion within 7 days of the clinical encounter. Our data captured the number of days between the date of the clinical encounter and the date the EPA was triggered by the learner (time from encounter to trigger [TET]). We ascertained the number of days between the trigger date and the date of EPA completion by the assessor (trigger to completion [TTC]). We gauged timeliness as the sum of TET and TTC. The 7-day measure for timeliness was based on the measure used by Tomiak and colleagues.[7]

Written feedback was labeled as "task oriented" if it commented on specific tasks or actions. Feedback

TABLE 1
Dimensions of Entrustable Professional Activity Feedback Analysis

Categories		Definition	Examples
Timely	Yes	$\leq$ 7 days	. . .
	No	$>$ 7 days	
Task oriented	Yes	Comments pertained to specific tasks or actions rather than learner attributes	"Ordered appropriate investigations and started initial treatment for patient."
	No		"Knowledge base is excellent. This was his first CTU block."
Actionable	Not actionable	No recommendations given	"Excellent job!"
	Semi-actionable	Recommendations that did not name a specific action or behavior but might identify a more general task or skillset to improve upon	"Continue to work on communication skills." "Continue to practice history and examination skills." "Read around your cases."
	Very actionable	Recommendations that targeted specific actions or behaviors	"Continue to encourage and guide open conversations regarding goals of care with patients and their families." "Read about signs/symptoms of opioid withdrawal and supportive management options."
Polarity[a]	Reinforcing	Comments endorsed or complimented the residents' performance	"Good work keeping track of the issues of a complex patient with a long hospital stay. You were very familiar with her history and what other services plans for the patient were."
	Corrective	Comments indicated problematic performance or need for improvement	"Had to direct on next steps and how to manage if patient became unstable."
	Mixed	Comments contained both positive and negative elements	"The conversations you had were great but could have been aided by a more directly suggested plan that DNR basic was most in keeping with the patient's wishes and overall clinical context."
	Neutral	Comments were left blank or described the clinical context without providing any comment on resident performance	"Complex medical presentation"

Abbreviations: EPA, entrustable professional activity; CTU, clinical teaching unit; DNR, do-not-resuscitate.
[a] Polarity was only assessed within the Comments section of the EPAs and did not take into consideration the Next Steps section.

was labeled as "very actionable" if recommendations gave targeted specific actions or behaviors, and "not actionable" if feedback gave no recommendations for development. Through our adjudication process, we identified comments that held value as feedback but fell in between our a priori definitions of very actionable and not actionable. We therefore thought it important to distinguish these comments and categorized them as "semi-actionable." Finally, narrative feedback was analyzed for polarity and was deemed as "reinforcing" if feedback complimented learners' performance, "corrective" if feedback identified problematic performance, "mixed" if comments contained both reinforcing and corrective elements, and "neutral" if no feedback was given or if comments did not address learner performance. Because feedback in the Next Steps section is meant to provide constructive recommendations, we based polarity only on feedback written in the Comments section. Please see TABLE 1 for a summary of these definitions and relevant examples.

EPAs were randomized and assigned to 2 among 4 independent reviewers (L.M., N.C., J.D., S.K.). All identifying data of the assessors and residents were removed by the program administrator prior to the study. Reviewers read the narratives within each EPA and assigned a code for each domain of quality feedback—timeliness (yes or no), task orientation (yes or no), actionability (very, semi-, or not actionable), and polarity (reinforcing, corrective, mixed, or neutral). Coding was completed in Microsoft Excel. Reviewers then met to discuss any disagreements in coding, which were resolved through consensus.

Statistical Analysis

Statistical analyses were completed using R 3.6.3 (The R Foundation). Interrater reliability for each domain

TABLE 2
Descriptive Statistics of Feedback Type Provided by Faculty and Resident Assessors

Feedback Domains	Total (n=1916)	Residents (n=1213), n (%)	Faculty (n=703), n (%)	χ^2 (*P* value)
Timely	926	561 (46)	337 (48)	0.5 (.48)
Task oriented	1679	1022 (84)	600 (85)	0.4 (.52)
Actionable				(.71)[a]
Not actionable	1386	857 (71)	490 (70)	
Semi-actionable	138	81 (7)	50 (7)	
Very actionable	457	275 (23)	163 (23)	
Polarity				(.007)[b]
Positive	1649	1033 (85)	564 (80)	
Mixed	79	35 (3)	40 (6)	
Negative	13	7 (1)	5 (1)	
Neutral	240	138 (11)	94 (13)	

[a] Cochran-Armitage test for trend (for ordinal data).
[b] Fisher's exact test.

within the feedback analyses was determined using Cohen's kappa (for nominal/binary variables) or weighted kappa (for ordinal variables). The level of agreement was interpreted as no ($\leq$0.20), minimal (0.21-0.39), weak (0.40-0.59), moderate (0.60-0.79), strong (0.80-0.90), or almost perfect (>0.90).[17]

Chi-square or Fisher's exact test (for nominal/binary variables) and Cochran-Armitage test for trend (for ordinal variables) were used to compare the type of feedback provided by faculty vs resident assessors, and to compare timely vs not timely feedback.

Lastly, to evaluate whether there was a timeliness by polarity interaction in the type of feedback received, a multivariable logistic regression model was used, including timely, polarity, and their interaction as covariates. Separate models were run for task oriented and actionable feedback as the outcome. A significant interaction term ($P<.05$) was indicative of an interaction. Given the sparse data for some categories, the polarity and actionable variables were dichotomized for the multivariable models.

As this study was conducted as part of a programmatic CQI initiative, ethics approval was not required according to local policy.[18]

Results

A total of 2471 EPAs were initiated by PGY-1 residents. Of these, 1981 (80%) were completed by assessors and were included in our analyses. Of these EPAs, 1213 (61%) were completed by senior resident or fellow physician supervisors, and the remainder were completed by attending physicians.

Interrater reliability of adjudicators was almost perfect for timeliness (κ=0.99), moderate for task orientation (κ=0.74), strong for actionability (κ=0.81), and moderate for polarity (κ=0.62). Senior resident assessors were all PGY-2 to PGY-5 residents. Analysis of the feedback showed that 47% (926 of 1981) of EPAs were timely. Median time for TET was 3 days (25th and 75th percentiles: 1 and 10 days). Median time for TTC was 2 days (25th and 75th percentiles: 0 and 10 days). Eighty-five percent (1679 of 1981) of feedback was task oriented. Regarding polarity of feedback, 83% (1649 of 1981) was reinforcing, 4% (79 of 1981) was mixed, and 12% (240 of 1981) was neutral.

Differences Between Resident and Faculty Assessors

Table 2 presents the type of feedback provided by residents and faculty advisors. Resident assessors were associated with providing more reinforcing feedback compared to faculty assessors (P=.007) based on Fisher's exact test. Residents and faculty did not differ with respect to timeliness ($\chi^2(1)$=0.5, P=.48) or task orientation ($\chi^2(1)$=0.4, P=.52). There was no difference between faculty and resident assessors in TTC. The Cochran-Armitage test for trend for ordinal data showed no difference in actionability of feedback between residents and faculty (z=0.37, P=.71).

Differences Between Timely and Not Timely Feedback

Table 3 presents the type of feedback provided, stratified by timeliness. The Cochran-Armitage test for trend showed that timely feedback was associated with feedback that was very actionable (z=3.11, P=.002). No difference in task orientation ($\chi^2(1)$=0.16, P=.69) or polarity ($\chi^2(3)$=1.76, P=.62)

TABLE 3
Type of Feedback Stratified by Timely Versus Not Timely

Feedback Domains	Not Timely (n=1055), n (%)	Timely (n=926), n (%)	χ^2 (*P* value)
Task oriented	891 (85)	788 (85)	0.2 (.69)
Actionable			(.002)[a]
Not actionable	766 (73)	620 (67)	
Semi-actionable	76 (7)	62 (7)	
Very actionable	213 (20)	244 (26)	
Polarity			1.8 (.62)
Positive	881 (84)	768 (83)	
Mixed	37 (4)	42 (5)	
Negative	6 (1)	7 (1)	
Neutral	131 (12)	109 (12)	

[a] Cochran-Armitage test for trend.

was identified between timely and not timely feedback. Lastly, the multivariable logistic regression did not identify a significant interaction between timeliness and polarity of feedback in terms of whether the feedback was task oriented (*P*=.16) or actionable (*P*=.40).

Discussion

Our study showed that, while most written feedback in EPAs was task oriented, fewer than half of the EPAs were completed in a timely manner. Moreover, timely feedback was correlated with greater actionability. Lastly, a greater percentage of mixed or corrective feedback was given by faculty, although faculty completed fewer EPAs compared to senior residents.[19] That only 47% (926 of 1981) of EPAs were completed in a timely manner suggests that this parameter can be improved. This finding, though concerning, is not surprising in that it likely reflects previously reported difficulties with allotting time to complete the forms themselves.[7] The recent national survey of Canadian residents by the Resident Doctors of Canada on the implementation of CBME reported that 32.9% of respondents perceived a lack of time in completing evaluations,[20] with written survey comments describing the time-consuming process of completing EPAs. Notably, 66.9% named evaluation fatigue as another barrier to CBME implementation.[20] Previous research has demonstrated how EPA completion for small programs, such as radiology, can add a significant administrative burden on those involved in the assessment process.[21] Thus, one potential way to alleviate this burden is to make the process itself more efficient by way of improved technology and dissemination process.[21] A survey of Canadian neurological surgeons showed that staff neurological surgeons were willing to complete an EPA if it took less than 3 minutes and if it was accessible through a mobile application.[22] One study of a mobile app for EPAs among psychiatry residents showed that the average time to complete an EPA via a mobile app was 76 seconds.[12] These improved technologies are promising avenues to increase the efficiency of EPA completion.

However, we note that the speed in completing EPAs may not correlate with the quality of feedback and may in fact compromise it. Therefore, an important yet more challenging approach to the issue of timeliness would be to reconsider the balance between the number of EPAs required and the quality of feedback/data each required EPA yields. Even though our data affirms the intuition that timely feedback correlates with actionable feedback, this does not account for the time it takes to complete multiple EPAs at a time.

Regardless of timeliness, the overall prevalence of actionable feedback was low in our study. This appears similar to the studies from Queen's University during initial phases of CBME implementation.[7,11] The finding of increased prevalence of actionable feedback over time may reflect a learning curve with CBME implementation. In the meantime, further faculty and resident development may be needed to develop their roles as coaches and assessors and to standardize the actionability of feedback.[23] Simple interventions such as the addition of prompts to elicit richer narrative feedback may also be effective.[5]

Improving the actionability of feedback—especially with corrective feedback—remains important because of evidence showing a lack of improvement in this area with CBME.[24] This is supported by our findings, as only 1% had corrective polarity. Moreover, this lack may reflect tensions assessors may have between their role as assessor and mentor/coach[23] as well as a prevailing culture of "failure to fail"[25,26] described in the literature.

Lastly, we note that a greater proportion of EPAs in our study were completed by senior residents compared to staff physicians. Some explanations for this may include the ability of senior residents to complete EPAs that require direct observation while on call. Residents may also have an increased level of comfort and trust[9,23] when asking for feedback and EPA completion. And while there is evidence to suggest that near-peer assessors provide similar ratings to staff physicians in low-stakes settings,[27] peer assessors also tend toward giving more favorable ratings[28]—a finding that we observed in our study. Thus, in the context of CBME, this tendency raises the question of whether senior residents can be relied upon as the prevailing drivers of completing EPA assessments and in gauging

the competence of their junior colleagues. Moreover, it also raises the question of whether the burden of assessment is disproportionately placed on residents rather than on attending physicians who were meant to give more feedback with CBME implementation.

Our study has several limitations. It was done in a single program and institution; therefore, our results may not be generalizable to other settings. Furthermore, assessing the quality of feedback remains subjective and context dependent. As reviewers, we interpreted EPA feedback apart from the original clinical context. We recognize that the quality of written feedback in EPAs does not necessarily reflect the feedback conversations that may have taken place during the respective clinical encounters. While our study was done within the context of IM, our methodology is translatable to other specialties that use EPAs to provide an approach to evaluating the quality of written feedback. Future studies should explore factors that contribute to the timeliness of EPA completion. Whether the proportion between faculty and resident assessors differs between institutions and specialties and the reasons why would also be important to explore further based on our study.

Conclusions

Overall, the written feedback in the EPAs we analyzed was task oriented but was neither timely nor actionable. Most of these EPAs were completed by senior residents rather than faculty.

References

1. Frank JR, Snell LS, ten Cate O, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638-645. doi:10.3109/0142159X.2010.501190
2. ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, Van Der Schaaf M. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach*. 2015;37(11):983-1002. doi:10.3109/0142159X.2015.1060308
3. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007;29(9-10):855-871. doi:10.1080/01421590701775453
4. Ginsburg S, van der Vleuten CPM, Eva KW, Lingard L. Cracking the code: residents' interpretations of written assessment comments. *Med Educ*. 2017;51(4):401-410. doi:10.1111/medu.13158
5. Hatala R, Sawatsky AP, Dudek N, Ginsburg S, Cook DA. Using In-Training Evaluation Report (ITER) qualitative comments to assess medical students and residents: a systematic review. *Acad Med*. 2017;92(6):868-879. doi:10.1097/ACM.0000000000001506
6. Branfield Day L, Miles A, Ginsburg S, Melvin L. Resident perceptions of assessment and feedback in competency-based medical education: a focus group study of one internal medicine residency program. *Acad Med*. 2020:95(11):1712-1717. doi:10.1097/ACM.0000000000003315
7. Tomiak A, Braund H, Egan R, et al. Exploring how the new entrustable professional activity assessment tools affect the quality of feedback given to medical oncology residents. *J Cancer Educ*. 2020;35(1):165-177. doi:10.1007/s13187-018-1456-z
8. Bing-You R, Hayes V, Varaklis K, Trowbridge R, Kemp H, McKelvy D. Feedback for learners in medical education: what is known? A scoping review. *Acad Med*. 2017;92(9):1346-1354. doi:10.1097/ACM.0000000000001578
9. 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;4(6):284-299. doi:10.1007/s40037-015-0231-7
10. Lockyer JM, Sargeant J, Richards SH, Campbell JL, Rivera LA. Multisource feedback and narrative comments: polarity, specificity, actionability, and CanMEDS roles. *J Contin Educ Health Prof*. 2018;38(1):32-40. doi:10.1097/CEH.0000000000000183
11. Tomiak A, Linford G, McDonald M, Willms J, Hammad N. Implementation of competency-based medical education in a Canadian medical oncology training program: a first year retrospective review. *J Cancer Educ*. 2022;37(3):852-856. doi:10.1007/s13187-020-01895-y
12. Young JQ, McClure M. Fast, easy, and good: assessing entrustable professional activities in psychiatry residents with a mobile app. *Acad Med*. 2020;95(10):1546-1549. doi:10.1097/ACM.0000000000003390
13. Martin L, Sibbald M, Brandt Vegas D, Russell D, Govaerts M. The impact of entrustment assessments on feedback and learning: trainee perspectives. *Med Educ*. 2020;54(4):328-336. doi:10.1111/medu.14047
14. Royal College of Physicians and Surgeons of Canada. Competence by Design Launch Schedule. Published 2021. Accessed January 24, 2021. https://www.royalcollege.ca/rcsite/cbd/cbd-implementation-e
15. Royal College of Physicians and Surgeons of Canada. Entrustable Professional Activities for Internal Medicine. Published 2019. Accessed October 14, 2022. http://www.royalcollege.ca/rcsite/documents/cbd/epa-observation-templates-e
16. Richards SH, Campbell JL, Walshaw E, Dickens A, Greco M. A multi-method analysis of free-text comments

from the UK general medical council colleague questionnaires. *Med Educ*. 2009;43(8):757-766. doi:10.1111/j.1365-2923.2009.03416.x
17. McHugh ML. Lessons in biostatistics interrater reliability: the kappa statistic. *Biochem Medica*. 2012;22(3):276-282.
18. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. Accessed November 4, 2022. https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html
19. Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: faculty interpretations of narrative evaluation comments. *Med Educ*. 2015;49(3):296-306. doi:10.1111/medu.12637
20. Resident Doctors of Canada. April 2021 National Resident Survey: Summary of Findings. Accessed November 4, 2022. https://residentdoctors.ca/wp-content/uploads/2021/10/RDoC-NRS-Summary-of-Findings-2021-R3.pdf
21. Cheung K, Rogoza C, Chung AD, Kwan BYM. Analyzing the administrative burden of competency based medical education. *Can Assoc Radiol J*. 2022;73(2):299-304. doi:10.1177/08465371211038963
22. Rabski JE, Saha A, Cusimano MD. Resident evaluations in the age of competency-based medical education: faculty perspectives on minimizing burdens. *J Neurosurg*. 2020:1-6. doi:10.3171/2020.7.JNS201688.
23. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76-85. doi:10.1111/medu.13645
24. Crawford L, Cofie N, McEwen L, Dagnone D, Taylor SW. Perceptions and barriers to competency-based education in Canadian postgraduate medical education. *J Eval Clin Pract*. 2020;26(4):1124-1131. doi:10.1111/jep.13371
25. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med*. 2005;80(suppl 10):84-87. doi:10.1097/00001888-200510001-00023
26. Gingerich A, Sebok-Syer SS, Larstone R, Watling CJ, Lingard L. Seeing but not believing: insights into the intractability of failure to fail. *Med Educ*. 2020;54(12):1148-1158. doi:10.1111/medu.14271
27. Khan R, Payne MWC, Chahine S. Peer assessment in the objective structured clinical examination: a scoping review. *Med Teach*. 2017;39(7):745-756. doi:10.1080/0142159X.2017.1309375
28. Moineau G, Power B, Pion AMJ, Wood TJ, Humphrey-Murto S. Comparison of student examiner to faculty examiner scoring and feedback in an OSCE. *Med Educ*. 2011;45(2):183-191. doi:10.1111/j.1365-2923.2010.03800.x

Lorenzo Madrazo, MD, is a PGY-4 General Internal Medicine Resident, Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada; **Jennifer DCruz, MD, MSc,** is a PGY-4 Respirology Resident, Division of Respirology, Queen's University, Kingston, Ontario, Canada; **Natasha Correa, MD,** is a PGY-4 Clinical Immunology and Allergy Resident, Division of Clinical Immunology and Allergy, University of Toronto, Toronto, Ontario, Canada; **Klajdi Puka, PhD,** is a Project Scientist, Institute for Mental Health Policy Research at the Centre for Addiction and Mental Health, Toronto, Ontario, Canada; and **Sheri-Lynn Kane, MD, FRCPC,** is Associate Professor and Previous Program Director, Internal Medicine, Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

Some of the content discussed in this article was previously presented at the virtual Canadian Conference on Medical Education, April 2021; the Western University Competency Based Medical Education Innovators Half Day, April 2021, London, Ontario, Canada; the Royal College 2021 Competency Based Medical Education Program Evaluation Summit, October 2021; and the virtual International Conference on Residency Education, October 20-22, 2021.

The authors would like to thank Ms. Ana Malbrecht for her work in preparing the data for analysis and to Dr. Christopher Watling for his input in conceptualizing this paper.

Corresponding author: Lorenzo Madrazo, MD, University of Ottawa, Ottawa, Ontario, Canada, lmadr080@uottawa.ca, Twitter @LorenzoMadrazo

Received March 9, 2022; revisions received July 18, 2022, and October 26, 2022; accepted October 31, 2022.