Exploring the Quality of Feedback in Entrustable Professional Activity Narratives Across 24 Residency Training Programs

Elizabeth A. Clement, MSc, MD
Anna Oswald, MD, MMEd
Soumyaditya Ghosh, PhD
Deena M. Hamza, PhD

ABSTRACT

Background Competency-based medical education (CBME) has been implemented in many residency training programs across Canada. A key component of CBME is documentation of frequent low-stakes workplace-based assessments to track trainee progression over time. Critically, the quality of narrative feedback is imperative for trainees to accumulate a body of evidence of their progress. Suboptimal narrative feedback will challenge accurate decision-making, such as promotion to the next stage of training.

Objective To explore the quality of documented feedback provided on workplace-based assessments by examining and scoring narrative comments using a published quality scoring framework.

Methods We employed a retrospective cohort secondary analysis of existing data using a sample of 25% of entrustable professional activity (EPA) observations from trainee portfolios from 24 programs in one institution in Canada from July 2019 to June 2020. Statistical analyses explore the variance of scores between programs (Kruskal-Wallis rank sum test) and potential associations between program size, CBME launch year, and medical versus surgical specialties (Spearman's rho).

Results Mean quality scores of 5681 narrative comments ranged from 2.0±1.2 to 3.4±1.4 out of 5 across programs. A significant and moderate difference in the quality of feedback across programs was identified (χ^2=321.38, P<.001, ε2=0.06). Smaller programs and those with an earlier launch year performed better (P<.001). No significant difference was found in quality score when comparing surgical/procedural and medical programs that transitioned to CBME in this institution (P=.65).

Conclusions This study illustrates the complexity of examining the quality of narrative comments provided to trainees through EPA assessments.

Introduction

Competency-based medical education (CBME) has been implemented in many residency training programs across Canada through a staged cohort approach, which is anticipated to be complete by 2027.[1] Transitioning to CBME is mandated across Canada; however, Pan-Canadian specialty committees with representation from all provinces meet to determine the best timing for all schools across Canada to launch each specific discipline. A key component of the CBME curriculum is the documentation of frequent low-stakes workplace-based assessments in addition to high-stakes assessments in order to track trainee progression over time.[2,3] Entrustable professional activities (EPAs) are designed to guide programs and trainees to relevant experiences by breaking down what it means to be a competent, safe, and effective practitioner into smaller, measurable, and achievable practice components, and they are one way that learner performance on workplace-based assessments is tracked.[4-6] Critically, the quality of narrative feedback is imperative for trainees to accumulate a body of evidence that tracks their progression to competence. Suboptimal narrative feedback used by committees responsible for reviewing a trainee's portfolio will challenge accurate decision-making, such as promotion to the next stage of training.[7]

Recent evidence suggests that the philosophy behind the development of a system of EPA-based observations is logical and promising; however, programs have encountered challenges with on-the-ground implementation.[8-11] Evidence suggests that challenges and variance in the quality of narrative feedback documented on EPA assessments may include preceptor-to-trainee ratio, size of program and specialty type (ie, surgical/procedural versus medical), level of supervision required to complete EPA observations, preceptor interpretation of the scale anchors, and perceived administrative burden of providing feedback.[12-21]

A noteworthy reflection about the role of EPAs as core building blocks of CBME is the uncertainty about translating the activities of CBME from theory into practice and speculating if they will lead to

DOI: http://dx.doi.org/10.4300/JGME-D-23-00210.1

anticipated outcomes. More specifically, there is uncertainty about the influence of EPA assessments on the competency of graduates entering unsupervised practice.[19] The purpose of this study is to explore the quality of documented feedback provided on workplace-based assessments at one institution in Canada by examining and scoring EPA narrative comments using a published quality scoring framework. Given previous literature outlining the challenges of providing feedback to trainees, this work measures the quality of feedback on EPAs in relation to CBME launch year, program size, and program type, and provides insight into the narrative comments that committees review when tasked with making decisions about a trainee's progression (or not) to the next stage of training.

KEY POINTS

What Is Known
Narrative feedback is a significant component of competency-based medical education, but the quality of such feedback has not been robustly reported at this scale and method to date.

What Is New
This quantitative study of entrustable professional activity (EPA) narrative comments from multiple specialties at a single institution in Canada showed that smaller programs and those with an earlier launch year were associated with higher-quality feedback; medical and surgical specialties had no difference in quality score.

Bottom Line
Understanding patterns where high-quality feedback can be found may help programs identify best practices as they look to improve their own EPA feedback quality.

Methods

Data Sources

To explore the quality of narrative feedback on EPA observation forms, we employed a retrospective cohort secondary analysis of existing data using de-identified EPA observations from trainee portfolios.

Data included in the analysis are from 24 programs in the July 2019 to June 2020 academic year at one institution in Canada. The institution is in an urban setting with more than 60 residency training programs. At the time of this study, there were more than 980 trainees enrolled in training. This study included only records from trainees that were in active Royal College of Physicians and Surgeons of Canada CBME programs (approximately 23 000 records) (TABLE 1). Program size ranged from 1 to 117 trainees, and all programs had been active in CBME for 1 to 3 years at the time of data collection.

EPA observation form adaptations at this institution include mandatory fields and prompts to guide documentation of quality feedback. For example, each EPA observation form contains one text box prompting the observer to describe what the trainee is doing well and why they should keep doing it, and another prompting the observer to describe something the trainee can do to improve their performance next time. The included narrative comments on EPA observation forms represent feedback provided to trainees from 24 specialty programs that have transitioned to CBME, representing a broad range of clinical disciplines, including medical, surgical, and diagnostic programs, as well as both primary specialty and subspecialty programs and, as such, also represent a large variety of EPA topics and contexts (see TABLE 1 for details). All narrative comments were exported into a Microsoft Excel (Excel version 16.61.1) document, and every fourth entry was scored, representing a selection of 25% of all EPA observations. Three raters (D.M.H., E.A.C., S.G.), 1 senior trainee and 2 PhD scientists, independently scored a subset of 10 narrative comments captured through EPA observations to establish initial agreement. An a priori minimum interrater agreement value of 80% was calculated using a commonly used approach of percentage agreement score (the number of agreements divided by the total number of EPA narrative assessed).[22] The raters met frequently to discuss quality scores, and discrepancies were resolved through consensus and group scorings of an additional 5 EPA narratives to calibrate a working understanding between team members. The senior scientist (D.M.H.) audited a subset of ratings (10%) to ensure rater agreement of the scoring (82% agreement score).

Quality Scoring Tool

A recently published quantitative framework was used to score the quality of narrative feedback. The Quality of Assessment of Learning (QuAL) score developed by Chan et al is specifically designed to rate short, workplace-based narrative comments on trainee performance, evaluating narrative comments based on 3 questions intended to capture evidence of quality feedback (FIGURE).[23] The QuAL score yields a numerical quality rating, with 5 being the highest score possible. A recent study explored the utility of the QuAL score based on narrative feedback provided on workplace-based assessments.[24]

Statistical Analyses

The Kruskal-Wallis rank sum test for nonparametric data was used to explore variance between programs, and epsilon squared ($\varepsilon 2$) was used to identify the magnitude of variance. Correlations to explore

TABLE 1
List of Included Medical and Surgical Programs, CBME Launch Year, and Number of Residents During the July 2019-June 2020 Academic Year

Program	CBME Launch Year	Number of Residents
Adult critical care medicine	2019	9
Adult gastroenterology	2019	10
Adult nephrology	2018	7
Anatomical pathology	2019	10
Anesthesiology[a]	2017	39
Cardiac surgery[a]	2019	15
Emergency medicine	2018	30
Forensic pathology	2018	1
General internal medicine	2019	15
General pathology	2019	8
Geriatric medicine	2019	4
Internal medicine	2019	106
Medical oncology	2018	3
Neurological surgery[a]	2019	12
Obstetrics and gynecology[a]	2019	34
Orthopaedic surgery[a]	2019	22
Otolaryngology[a]	2017	17
Pediatric critical care medicine	2019	5
Pediatric nephrology	2018	1
Radiation oncology	2019	11
Rheumatology	2019	5
Surgical foundations[a]	2018	117
Urology[a]	2018	17

Abbreviation: CBME, competency-based medical education.
[a] Surgical/procedural specialties as conceptualized by the authors.
Note: The number of residents is based on administrative data provided to the study team and does not account for changes in the number of residents during the course of the academic year.

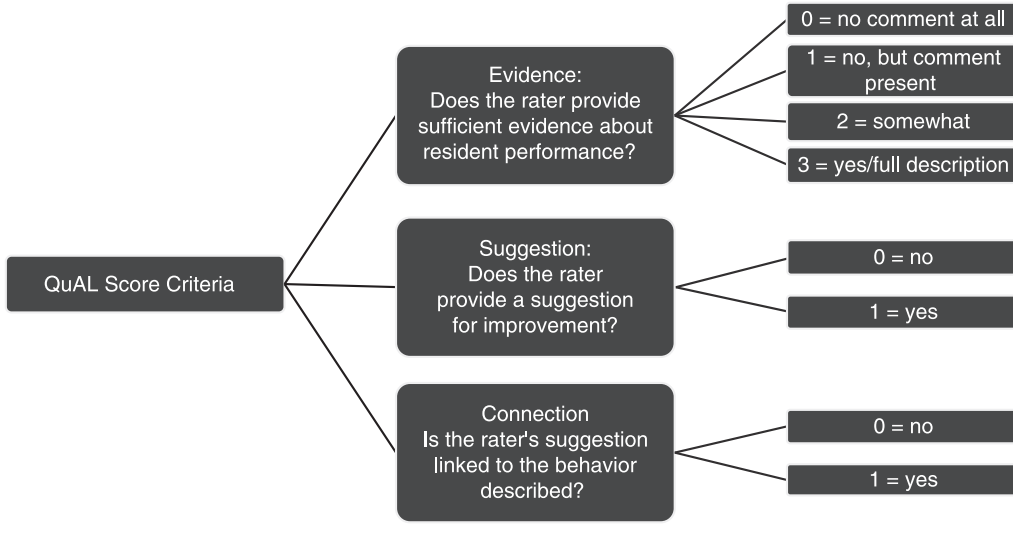

FIGURE
Criteria Used to Assess the Quality of Entrustable Professional Activity Observation Narratives Using the Quality of Assessment of Learning (QuAL) Score

TABLE 2
Examples of Low, Moderate, and High-Quality QuAL Scores on Narrative Feedback in EPA Observation Forms

EPA Observation Narrative Comment	QuAL Score Criteria With Examples From EPA Observation Narrative Comments			
	Evidence (/3)	Suggestion (/1)	Connection (/1)	Total QuAL Score
None. Mandatory comment fields bypassed by entering a period	0: No comment made	0: No suggestion	0: No connection	0
"Procedure performed well"	1: Comment is made, no behavior description or patient context	0: No suggestion	0: No connection	1
"Doing a thorough job; Keep working on breastfeeding advice"	1: Comment made, no behavior description or patient context	1: Suggestion made	0: Connection unclear	2
"Complex bronch, took careful time and attention to recognize unique anatomy, and also to balance risk/benefit of pursuing further intervention (suction or leave clot alone)"	3: Detailed description of resident behavior and patient context	0: No suggestion	0: No connection	3
"Well thought through with a contingency/back up plan. Recognizes the nuances, and not just the 'numbers' for successful extubation. Suggest additional teaching around the case to help the residents understand some of the nuances and pitfalls to extubation."	2: Detailed description of resident behavior, but no context provided	1: Suggestion included	1: Suggestion is connected to evidence	4
"Identifying most concerning diagnosis and treated appropriately in context of patient risk factors (use of plavix due to high bleeding risk). Considered other ddx, and created plan (discuss with heme onc re chemo, considering v/q scan). Discussed care with family regarding decision for angio; make plan for what to do with other rate slowing agents if placing on new medication."	3: Detailed description of resident behavior and patient context	1: Suggestion included	1: Suggestion is connected to evidence	5

Abbreviations: QuAL, Quality of Assessment of Learning; EPA, entrustable professional activity.

the association between quality scores and program size, CBME launch year, and medical versus surgical/procedural specialities were calculated using Spearman's rho (ρ).

This study was approved by the institution's Research & Ethics Board.

Results

A total of 5681 documented EPA observations (every fourth narrative comment scored; 25% selected from each program) from 24 CBME programs were examined using the QuAL score. TABLE 2 illustrates examples of narrative feedback with QuAL scores.

Mean QuAL scores, as calculated within programs, ranged from 2.0±1.2 to 3.4±1.4 (± represent margin of error, 95% confidence intervals) out of 5 across programs. Overall, a significant and moderate difference in the quality of feedback provided to trainees across programs was identified (χ^2=321.38, P<.001, ε2=0.06). The quality of feedback is significantly associated with program size as well as launch year; smaller programs and those with earlier launch year performed better (P<.001). This association, however, is very weak (ρ=0.08, 0.05 respectively). Further, there was no significant difference in performance when comparing surgical and medical specialty programs (P=.65, ρ=-0.08).

Discussion

Results from scoring 5681 de-identified EPA observation records from 24 CBME programs indicate

that the quality of documented feedback on EPA observation forms demonstrate a wide range of quality scores within and between programs, yet scores overall are on the lower to middle range based on the QuAL scoring tool (2.0±1.2 to 3.4±1.4 out of 5) and could be improved. These results build on the published literature about the implementation of CBME in which the importance of quality documented feedback is reported by both trainees and clinical teachers, yet the actual documentation of consistent quality feedback can be challenging.[16,19,25,26] Considering the critical role documented feedback plays in the CBME training model, such as supporting learner growth over time and facilitating group decision-making about learner progression through training, these results add rigour to the existing evidence suggesting a discrepancy between the grand aspirations behind the design and the subsequent local implementation of this model that may challenge long-lasting change of CBME activities.[27]

Despite the results of this study, it is important to note that the quality of documented feedback is not the sole indicator of the quality of the feedback exchange between trainees and clinical teachers. Recent literature suggests that, in some cases, trainees prefer the informal feedback conversation that occurs prior to the completion of an EPA observation form, and that this feedback tends to be richer than what is documented.[16,28] Faculty development initiatives have been created to support the new practices and the shift to a culture of CBME. These often aim to develop supervisors' skills in providing rich verbal coaching conversations between trainees and clinical teachers while also supporting these supervisors' ability to sufficiently document these encounters to support the requirements for progression recommendations by competence committees.[18,29,30] Other areas to consider include (1) the administrative burden of EPA assessment on both preceptors and trainees, specific to balancing timeliness of feedback with quality of content; (2) similarities and differences in how quality feedback is conceptualized by preceptors and trainees; and (3) subsequent utility of the content provided in different contexts (eg, feedback in order to make progress decisions; feedback to foster actionable growth of the trainee).[19,31,32]

While the QuAL scoring tool demonstrates validity evidence for assessing the quality of brief workplace-based assessments, like EPA observations, the interpretation of the criteria by raters is subjective and may not align with constructs of helpful or useful feedback in all contexts. As such, a different group of raters could score aspects of the QuAL, like the evidence criterion, differently. There is also a limitation to the commonly used percentage-agreement method to calculating interrater reliability, as the value does not account for agreements that may be due to chance[22,33]; however, chance agreements are more likely to occur in binary yes/no decisions. The authors acknowledge the subjective nature of assessing the quality of narrative comments and have detailed their processes for transparency.

Another limitation is that EPA observation narrative comments included in this study are from one institution. Additionally, narrative comments examined include those collected during the COVID-19 pandemic. For privacy and confidentiality, timestamps of when EPA narrative comments were logged were removed from the dataset by external data stewards, and we are unable to perform an analysis that accounts for changes during this time. Finally, this methodology cannot determine the reasons why quality of narrative feedback may be lower than desired, such as electronic system barriers, skills deficits in the feedback providers, availability of resources, and potential administrative burden. However, assessing the quality of documented feedback provides valuable information about the real-world experiences of group decision-makers, as they are responsible for making progress recommendations about trainees and subsequent targets for improvement using EPA narratives as a source of evidence.

Future studies that include demographic information about learners and preceptors, and that explore the contexts and factors contributing to the quality of documented workplace-based assessments are necessary. These could potentially be multi-institutional studies to promote the transferability of findings or qualitative studies to better understand the reasons why quality was lower than desired. Further, it would be helpful to examine how efforts to improve supervisors' narrative comments affect subsequent quality scores over time, as well as what influence external factors, such as COVID-19, have on the quality of narrative comments. It would also be of practical importance to see the potential of automated narrative comment quality rating systems and their comparability with the QuAL score to allow for incorporation of live quality reporting in electronic portfolio systems. Additional studies exploring ways to balance the challenges of implementing the practices and activities of CBME and the ambitions for the change initiative are needed, along with mobilizing these findings into practice.

Conclusions

The results suggest that smaller programs and those that have implemented CBME training for a longer time have somewhat higher-quality narrative comments

on EPAs, with no difference observed between medical and surgical/procedural specialities.

References

1. Royal College of Physicians and Surgeons of Canada. CBD start, launch and exam schedule 2022. Accessed November 27, 2023. https://www.royalcollege.ca/content/dam/documents/accreditation/competence-by-design/directory/cbd-rollout-schedule-e.html
2. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676-682. doi:10.3109/0142159X.2010.500704
3. Lockyer J, Carraccio C, Chan MK, et al. Core principles of assessment in competency-based medical education. *Med Teach*. 2017;39(6):609-616. doi:10.1080/0142159X.2017.1315082
4. ten Cate O. Nuts and bolts of entrustable professional activities. *J Grad Med Educ*. 2013;5(1):157-158. doi:10.4300/JGME-D-12-00380.1
5. Shorey S, Lau TC, Lau ST, Ang E. Entrustable professional activities in health care education: a scoping review. *Med Educ*. 2019;53(8):766-777. doi:10.1111/medu.13879
6. ten Cate O, Taylor DR. The recommended description of an entrustable professional activity: AMEE guide no. 140. *Med Teach*. 2021;43(10):1106-1114. doi:10.1080/0142159X.2020.1838465
7. Norman G, Norcini J, Bordage G. Competency-based education: milestones or millstones? *J Grad Med Educ*. 2014;6(1):1-6. doi:10.4300/JGME-D-13-00445.1
8. Bentley H, Darras KE, Forster BB, Sedlic A, Hague CJ. Review of challenges to the implementation of competence by design in post-graduate medical education: what can diagnostic radiology learn from the experience of other specialty disciplines? *Acad Radiol*. 2022;29(12):1887-1896. doi:10.1016/j.acra.2021.11.025
9. Carraccio C, Englander R, Van Melle E, et al. Advancing competency-based medical education: a charter for clinician-educators. *Acad Med*. 2016;91(5):645-649. doi:10.1097/ACM.0000000000001048
10. Dagnone JD, Chan MK, Meschino D, et al. Living in a world of change: bridging the gap from competency-based medical education theory to practice in Canada. *Acad Med*. 2020;95(11):1643-1646. doi:10.1097/ACM.0000000000003216
11. Touchie C, ten Cate O. The promise, perils, problems and progress of competency-based medical education. *Med Educ*. 2016;50(1):93-100. doi:10.1111/medu.12839
12. Royal College of Physician and Surgeons of Canada. CBD Program Evaluation Recommendations Report: 2019. https://www.collegeroyal.ca/
13. Arnstead N, Campisi P, Takahashi SG, et al. Feedback frequency in competence by design: a quality improvement initiative. *J Grad Med Educ*. 2020;12(1):46-50. doi:10.4300/JGME-D-19-00358.1
14. El-Haddad C, Damodaran A, McNeil HP, Hu W. The ABCs of entrustable professional activities: an overview of 'entrustable professional activities' in medical education. *Intern Med J*. 2016;46(9):1006-1010. doi:10.1111/imj.12914
15. O'Dowd E, Lydon S, O'Connor P, Madden C, Byrne D. A systematic review of 7 years of research on entrustable professional activities in graduate medical education, 2011-2018. *Med Educ*. 2019;53(3):234-249. doi:10.1111/medu.13792
16. 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
17. Upadhyaya S, Rashid M, Davila-Cervantes A, Oswald A. Exploring resident perceptions of initial competency based medical education implementation. *Can Med Educ J*. 2021;12(2):e42-e56. doi:10.36834/cmej.70943
18. van Loon KA, Driessen EW, Teunissen PW, Scheele F. Experiences with EPAs, potential benefits and pitfalls. *Med Teach*. 2014;36(8):698-702. doi:10.3109/0142159X.2014.909588
19. Hamza DM, Hauer KE, Oswald A, et al. Making sense of competency-based medical education (CBME) literary conversations: a BEME scoping review: BEME guide no. 78. *Med Teach*. 2023;45(8):802-815. doi:10.1080/0142159X.2023.2168525
20. Postmes L, Tammer F, Posthumus I, Wijnen-Meijer M, van der Schaaf M, ten Cate O. EPA-based assessment: clinical teachers' challenges when transitioning to a prospective entrustment-supervision scale. *Med Teach*. 2021;43(4):404-410. doi:10.1080/0142159X.2020.1853688
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. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276-282.
23. Chan TM, Sebok-Syer SS, Sampson C, Monteiro S. The Quality of Assessment of Learning (Qual) score: validity evidence for a scoring system aimed at rating short, workplace-based comments on trainee performance. *Teach Learn Med*. 2020;32(3):319-329. doi:10.1080/10401334.2019.1708365
24. Woods R, Singh S, Thoma B, et al. Validity evidence for the Quality of Assessment for Learning score: a quality metric for supervisor comments in competency based medical education. *Can Med Educ J*. 2022;13(6):19-35. doi:10.36834/cmej.74860

25. 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
26. Marcotte L, Egan R, Soleas E, Dalgarno N, Norris M, Smith C. Assessing the quality of feedback to general internal medicine residents in a competency-based environment. *Can Med Educ J*. 2019;10(4):e32-e47.
27. Hamza DM, Regehr G. Eco-normalization: evaluating the longevity of an innovation in context. *Acad Med*. 2021;96(suppl 11):48-53. doi:10.1097/ACM.0000000000004318
28. 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
29. Bray MJ, Bradley EB, Martindale JR, Gusic ME. Implementing systematic faculty development to support an epa-based program of assessment: strategies, outcomes, and lessons learned. *Teach Learn Med*. 2021;33(4):434-44. doi:10.1080/10401334.2020.1857256
30. ten Cate O. Entrustment as assessment: recognizing the ability, the right, and the duty to act. *J Grad Med Educ*. 2016;8(2):261-262. doi:10.4300/JGME-D-16-00097.1
31. Ott MC, Pack R, Cristancho S, Chin M, Van Koughnett JA, Ott M. "The most crushing thing": understanding resident assessment burden in a competency-based curriculum. *J Grad Med Educ*. 2022;14(5):583-592. doi:10.4300/JGME-D-22-00050.1
32. Madrazo L, DCruz J, Correa N, Puka K, Kane SL. Evaluating the quality of written feedback within entrustable professional activities in an internal medicine cohort. *J Grad Med Educ*. 2023;15(1):74-80. doi:10.4300/JGME-D-22-00222.1
33. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34. doi:10.20982/tqmp.08.1.p023

Elizabeth A. Clement, MSc, MD, is a General and Colorectal Surgeon, Division of General Surgery, Department of Surgery, St. Paul's Hospital, University of British Columbia, Vancouver, British Columbia, Canada, and Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada; **Anna Oswald, MD, MMEd,** is a Rheumatologist, Director of CBME Implementation, and Clinicia Educator, Division of Rheumatology, Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada; **Soumyaditya Ghosh, PhD,** is a Research Associate, Postgraduate Medical Education, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada; and **Deena M. Hamza, PhD,** is Innovations and Health Professions Education Scientist, Postgraduate Medical Education, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada.

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

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

This work was previously presented at Resident Research Day, University of Alberta, April 22, 2021, Edmonton, Alberta, Canada, and the virtual International Conference on Residency Education, October 22, 2021.

Corresponding Author: Deena M. Hamza, PhD, University of Alberta, Edmonton, Alberta, Canada, Dh9@ualberta.ca, X @DrDeenaMHamza

Received March 22, 2023; revisions received July 15, 2023, and November 17, 2023; accepted November 27, 2023.