Evaluation of a National Competency-Based Assessment System in Emergency Medicine: A CanDREAM Study

Brent Thoma, MD, MA, MSc, FRCPC, DRCPSC
Andrew K. Hall, MD, FRCPC, MMEd
Kevin Clark, MD, MMEd, FRCPC
Nazanin Meshkat MD, FRCPC, MHSc
Warren J. Cheung, MD, MMEd, FRCPC
Pierre Desaulniers, MD, CSPQ
Cheryl ffrench, MD, FRCPC
Allison Meiwald, BN, MD, MEd, FRCPC
Christine Meyers, MD, FRCPC
Catherine Patocka, MDCM, MHPE, FRCPC
Lorri Beatty, MD, FRCPC
Teresa M. Chan, MD, MHPE, FRCPC, DRCPSC

ABSTRACT

Background In 2018, Canadian postgraduate emergency medicine (EM) programs began implementing a competency-based medical education (CBME) assessment program. Studies evaluating these programs have focused on broad outcomes using data from national bodies and lack data to support program-specific improvement.

Objective We evaluated the implementation of a CBME assessment program within and across programs to identify successes and opportunities for improvement at the local and national levels.

Methods Program-level data from the 2018 resident cohort were amalgamated and analyzed. The number of entrustable professional activity (EPA) assessments (overall and for each EPA) and the timing of resident promotion through program stages were compared between programs and to the guidelines provided by the national EM specialty committee. Total EPA observations from each program were correlated with the number of EM and pediatric EM rotations.

Results Data from 15 of 17 (88%) programs containing 9842 EPA observations from 68 of 77 (88%) EM residents in the 2018 cohort were analyzed. Average numbers of EPAs observed per resident in each program varied from 92.5 to 229.6, correlating with the number of blocks spent on EM and pediatric EM ($r = 0.83$, $P < .001$). Relative to the specialty committee's guidelines, residents were promoted later than expected (eg, one-third of residents had a 2-month delay to promotion from the first to second stage) and with fewer EPA observations than suggested.

Conclusions There was demonstrable variation in EPA-based assessment numbers and promotion timelines between programs and with national guidelines.

Introduction

As competency-based medical education (CBME) is being implemented around the world,[1] it is also being evaluated to quantify its impact and support its improvement. Evaluation studies published to date focus on broad outcomes using data from national bodies such as the Accreditation Council for Graduate Medical Education (ACGME)[2–4] or emphasize the outcomes from local[5–9] and regional[10,11] implementation. While national analyses can inform the evolution of an overall assessment program, they provide insufficient data to support *program-specific* improvement.[2–4] Conversely, local or regional initiatives reveal insights within their context, but it is unclear whether they represent a broader systemic challenge.[2,10,11] Neither type of database is able to detect variability or fidelity of implementation[7,12,13] across individual programs, an essential first step in evaluating higher-level educational and clinical outcomes.[14] Regardless of the specialty, this is a problem that any program must face when implementing CBME.

Emergency medicine (EM) residency programs accredited by the Royal College of Physicians and Surgeons of Canada (RCPSC) officially implemented their CBME assessment program for the cohort of residents beginning postgraduate training in July 2018 (the 2018 cohort).[15] This assessment program consists of 28 entrustable professional activities (EPAs) assessed on a 5-point entrustment scale[16,17] that are organized sequentially into 4 stages (Transition to Discipline, Foundations of Discipline, Core of Discipline, and Transition to Practice) spread across 5 years of training (TABLE 1), all of which were predetermined centrally by the RCPSC EM specialty committee.[15] The specialty committee also suggested a target number of assessments for each EPA. These

DOI: http://dx.doi.org/10.4300/JGME-D-19-00803.1

Editor's Note: The online version of this article contains a data extraction spreadsheet completed by each of the 15 participating Canadian emergency medicine residency training programs and the O-SCORE and University of Toronto entrustment scales used to assess each entrustable professional activity.

targets were determined by the specialty committee members.[18] While the EM CBME assessment program has a consistent design across sites, the roll-out of the program was site-specific.

We evaluated the short-term outcomes of the national implementation of this assessment program for Canadian RCPSC EM training programs through the creation of a specialty-specific database of program-level assessment data.[14] This evaluation aimed to identify successes and opportunities for improvement at local and national levels, investigate the fidelity of implementation[13,19] of the new program of assessment, evaluate the variability of implementation between training programs and the fidelity of the implementation relative to the national design, and present analyses that support the improvement of local programs and the national assessment program.

What was known and gap
Studies evaluating competency-based medical education (CBME) assessment for postgraduate emergency medicine programs in Canada have focused on broad outcomes using data from national bodies and lack data to support program-specific improvement.

What is new
An evaluation of the implementation of a CBME assessment program within and across programs to identify successes and opportunities for improvement at the local and national levels.

Limitations
The study includes only the initial quantitative data for the first year of our implementation. The small sample size reduces generalizability.

Bottom line
Involving and engaging program-level educational leaders to collect and aggregate data can yield unique analytics that are useful to both local and national stakeholders and leaders.

Methods

The RCPSC has directed the implementation of CBME[20] sequentially by specialty in concert with national specialty committees.[15] As required by the RCPSC for each specialty, the EM specialty committee was founded in the early 1980s when EM was established as a training program. It consists of an executive (chair, vice-chair), representatives from 5 geographic constituencies across Canada, and the program directors from all institutions.

As part of the CBME rollout, each program established a competency committee charged with making decisions regarding promotion between stages by aggregating, analyzing, and reviewing each residents' assessment data. The RCPSC competency committees are structurally similar to the clinical competency committees used by the ACGME.[21–23] The methods the committees used to arrive at their decisions are idiosyncratic and locally derived.[24]

Enrollment of Programs

The program director or CBME faculty lead of each of the 14 Canadian institutions that host specialty EM residency programs was contacted and asked to participate. Representatives from 12 institutions overseeing 15 of the 17 programs agreed to participate. The 4 University of British Columbia's training sites were considered independent residency programs for the purpose of the analyses because their schedules differ, and their promotion decisions are conducted by independent competence committees.

Data Collection

Deidentified EPA assessment data was collected for residents who began residency in the 2018 cohort. We designed a 3-tab data extraction spreadsheet (provided as online supplemental material) to collect CBME data and relevant program characteristics from each program lead. The first tab contained the details of EPA observations (the number of observations of each EPA that occurred at each level of the 5-point entrustment scale[16,17]) from the included residents that were collected between July 1, 2018, and June 30, 2019. The second tab amalgamated data from the first tab into program-level metrics, including the total and mean (standard deviation [SD]) number of each EPA observed at each level of the entrustment scale. The third tab contained program characteristics, including the number of eligible residents in the 2018 cohort, the number of EM and pediatric EM training 4-week blocks within the first year, the number of shifts per EM training block, the number of residents in each stage of training as of the first day of each month (July 1, 2018, to July 1, 2019), and any additional information that each program lead felt was important to contextualize the data.

Ethics and Confidentiality

Our protocol was submitted to the Research Ethics Board at 12 institutions and deemed exempt by each as a program evaluation activity under article 2.5 of the national Tri-Council Policy Statement.[25] All data were deidentified by home program, and only program-level data were analyzed. One contact (K.C.) extracted data from all 4 UBC programs.

Data Analysis

Stage-specific analyses and visualizations excluded the final stage of residency (Transition to Practice)

TABLE 1
List of Entrustable Professional Activities (EPAs) and Suggested Number of Observations for Each and Stage Length

EPA Code	EPA Text	Suggested No. of Observations
Transition to Discipline (TD): ~3 months		
TD1	Recognizing the unstable/critically ill patient, mobilizing the health care team and supervisor, and initiating basic life support	10
TD2	Performing and documenting focused histories and physical examinations, and providing preliminary management of cardinal emergency department presentations	20
TD3	Facilitating communication of information between a patient in the emergency department, caregivers, and members of the health care team to organize care and disposition of the patient	10
Foundations of Discipline (F): ~9 months		
F1	Initiating and assisting in resuscitation of critically ill patients	15
F2	Assessing and managing patients with uncomplicated urgent and non-urgent emergency department presentations	30
F3	Contributing to the shared work of the emergency department health care team to achieve high-quality, efficient, and safe patient care	10
F4	Performing basic procedures	25
Core of Discipline (C): < 3 years		
C1	Resuscitating and coordinating care for critically ill patients	40
C2	Resuscitating and coordinating care for critically injured trauma patients	25
C3	Providing airway management and ventilation	20
C4	Providing emergency sedation and systemic analgesia for diagnostic and therapeutic procedures	20
C5	Identifying and managing patients with emergent medical or surgical conditions	40
C6	Diagnosing and managing patients with complicated urgent and non-urgent patient presentations	40
C7	Managing urgent and emergent presentations for pregnant and postpartum patients	15
C8	Managing patients with acute toxic ingestion or exposure	15
C9	Managing patients with emergency mental health conditions or behavioral emergencies	15
C10	Managing and supporting patients in situational crisis to access health care and community resources	5
C11	Recognizing and managing patients who are at risk of exposure to, or who have experienced violence and/or neglect	5
C12	Liaising with prehospital emergency medical services	5
C13	Performing advanced procedures	25
C14	Performing and interpreting point-of-care ultrasound to guide patient management	50
C15	Providing end-of-life care for a patient	5
Transition to Practice (TP): ~1+ year(s)		
TP1	Managing the emergency department to optimize patient care and department flow	25
TP2	Teaching and supervising the learning of trainees and other health care professionals	15
TP3	Managing complex interpersonal interactions that arise during the course of patient care	5
TP4	Providing expert emergency medicine consultation to physicians or other health care providers	5
TP5	Coordinating and collaborating with health care professional colleagues to safely transition the care of patients, including handover and facilitating inter-institution transport	10
TP6	Dealing with uncertainty when managing patients with ambiguous presentations	5

TABLE 2
Characteristics of Participating Programs' and 2018 Cohort of First-Year Residents Arranged by Number of Residents

Program	No. of Residents	Shifts per Block of EM	EM Blocks	Pediatric EM Blocks	Total No. of EM Blocks
Dalhousie University	2	16.0	6.0	2.0	8.0
McGill University	4	15.8	6.5	1.0	7.5
McMaster University	8	15.2	5.0	1.0	6.0
Queen's University	4	14.0	8.0	1.0	9.0
Université de Montréal	4	18.0	8.5	1.0	9.5
University of British Columbia (Fraser)	2	16.0	4.0	1.0	5.0
University of British Columbia (Interior)	2	16.0	5.0	1.0	6.0
University of British Columbia (Island)	2	16.0	4.0	1.0	5.0
University of British Columbia (Vancouver)	6	16.0	4.0	2.0	6.0
University of Calgary	4	16.0	6.0	2.0	8.0
University of Manitoba	4	18.0	7.0	1.0	8.0
University of Ottawa	9	15.0	6.5	2.0	8.5
University of Saskatchewan	3	14.0	8.5	2.0	10.5
University of Toronto	10	16.0	7.0	1.0	8.0
Western University	4	17.8	6.5	2.0	8.5
Average (SD)	4.5 (2.6)	16.0 (1.2)	6.2 (1.5)	1.4 (0.5)	7.6 (1.6)

Abbreviation: EM, emergency medicine.
Note: Some rotations were combination rotations with other smaller blocks (eg, prehospital care, point-of-care ultrasound) and therefore were assigned 0.5 a rotation. Also, some schools had a variable number of shifts over the 13 blocks of a year for their EM blocks. As such, some schools had an average number of shifts per block that is reported above.

because it contained minimal data. Descriptive statistics were calculated using Microsoft Excel 14.7.0 (Microsoft Corp, Albany, NY) and SPSS Statistics 25.0 (IBM Corp, Armonk, NY). Graphs were created using Microsoft Excel 16.0.1 (Microsoft Corp, Albany, NY). The relationship between the average number of EPA observations per resident within each program and the number of training blocks spent on EM and pediatric EM training blocks was evaluated with a Pearson's correlation.

Results

Descriptive Data on Program Sites

Data from 15 of 17 (88%) RCPSC EM programs containing 68 of the 77 (88%) residents in the 2018 cohort were analyzed. Combined, the residents received 9842 EPA observations in the study period. TABLE 2 outlines the characteristics of each of the programs, which demonstrated variability in the mean number of EM blocks (6.2, SD 1.5), pediatric EM blocks (1.4, SD 0.5), and shifts per EM block (16.0, SD 1.2).

Program-Level Data Analysis

FIGURE 1 demonstrates the variability in the average number of EPA observations across the 15 programs with a range of 92.5 to 229.5 EPA observations per resident. The variability in the average number of EPA observations completed within each stage is also represented within each bar of this FIGURE. The average (SD) values across the 15 programs were 45.6 (SD 8.7) Transition to Discipline EPAs observations per resident, 70.4 (SD 25.8) Foundations of Discipline EPA observations per resident, and 29 (SD 23.2) Core of Discipline EPA observations per resident.

FIGURE 2 is a stack chart representing the proportion of the 68-resident cohort in each stage of training on the first day of each month of the year. Although the specialty committee estimated that the Transition to Discipline stage would take approximately 3 months, one-third of residents were not promoted to the Foundations of

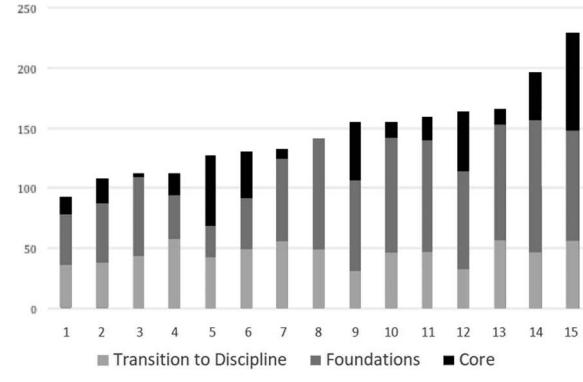

FIGURE 1
Modified Stack Chart Demonstrating Average Number of EPA Observations per Resident Within Each Program (Total and Each Stage of Training)

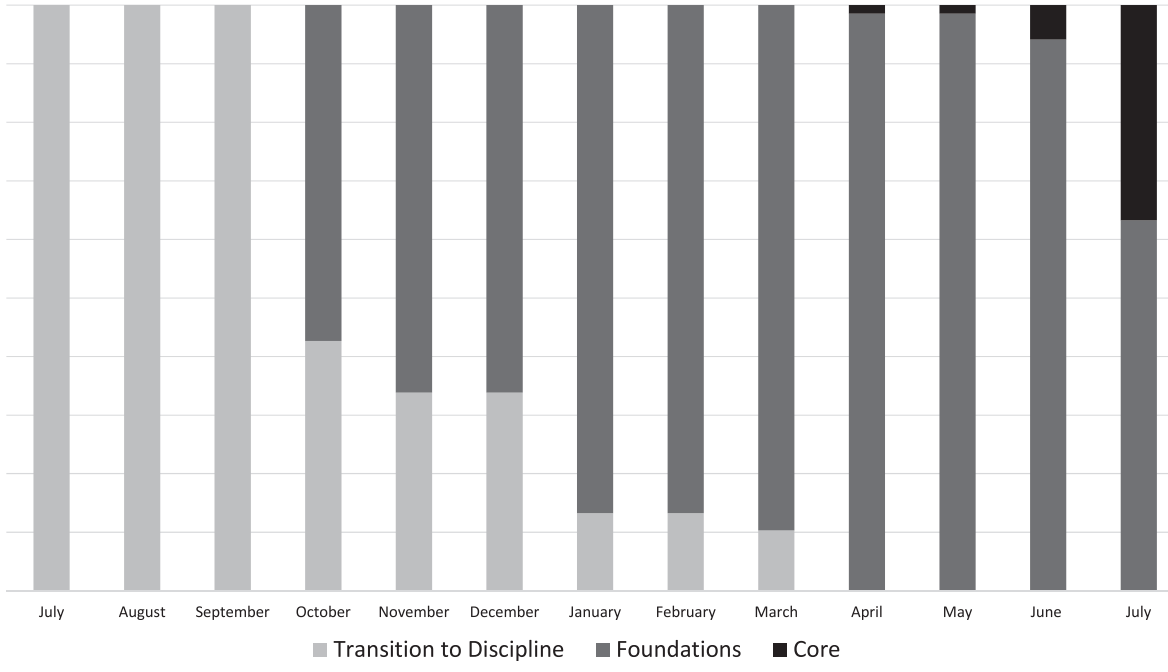

FIGURE 2
Stack Chart Demonstrating Percentage of First-Year Residents in Each Stage on First Day of Each Month (July 1, 2018–July 1, 2019)

Discipline stage for at least 5 months. Similarly, it was anticipated that the Foundations of Discipline stage would last until the end of the first year of residency, but over 60% of residents were not promoted to the Core of Discipline stage by the end of the year.

Aggregate Performance Analytic

FIGURE 3 outlines the average number of EPA observations per resident within each stage of training compared to the provided guidelines. All residents were promoted to the Foundations of Discipline

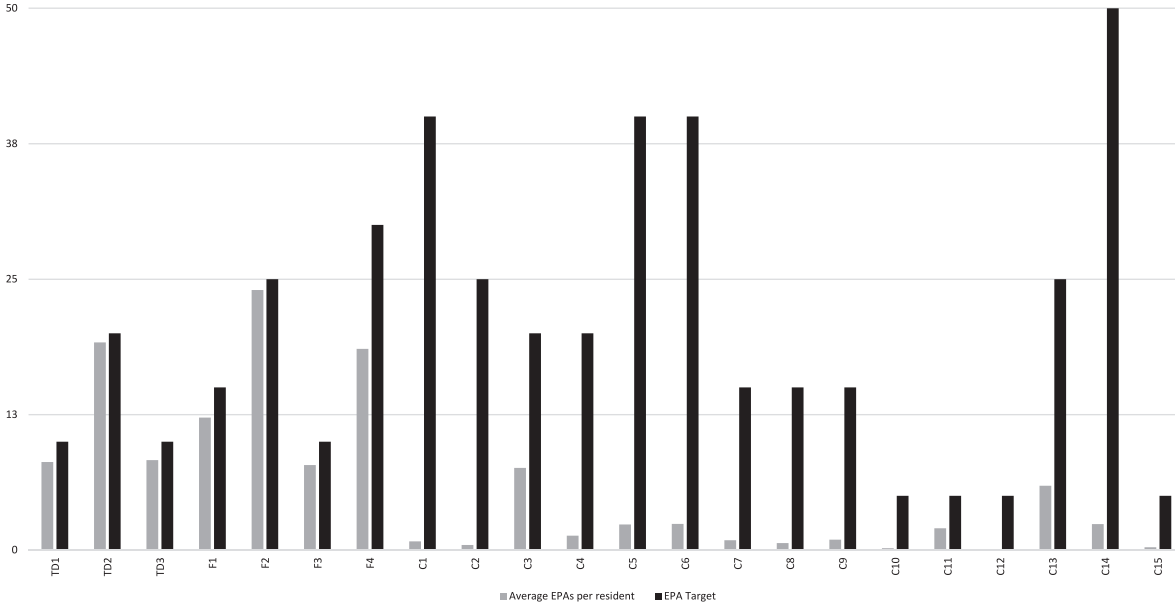

FIGURE 3
Bar Chart Demonstrating Average Number of EPAs Observed per Resident After 1 Year of Assessment Relative to Targeted Number Required for Promotion to Next Stage
Note: Descriptions of each EPA are shown in TABLE 1.

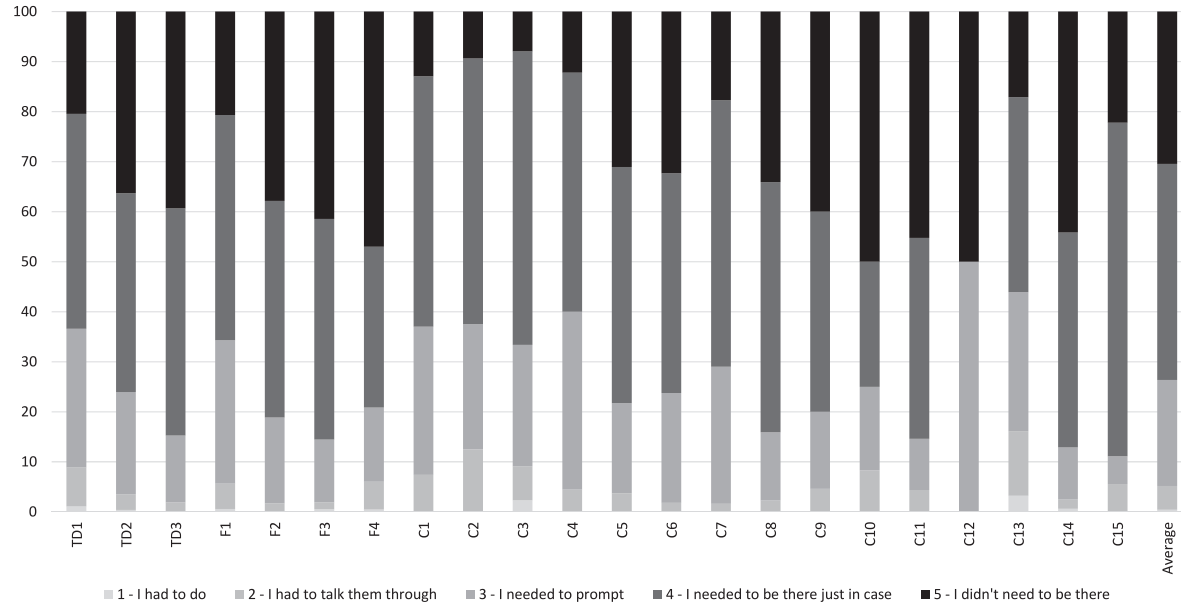

FIGURE 4
Stack Chart Demonstrating Percentage of Observations of Each EPA Rated at Each Level of Entrustment on the Ottawa Score

stage, and the average number of observations of the Transition to Discipline EPAs was less than the number recommended by the specialty committee. The average number of EPA observations prior to promotion to the Core of Discipline could not be assessed as most residents did not enter this stage before the end of the data collection period.

As individual resident assessment data were not obtained, we were unable to report traditional learning curves for individual EPAs. In lieu of a learning curve, FIGURE 4 represents the relative difficulty of each of the EPAs by presenting the proportion of all assessments that were scored at each level of the 5-point entrustment scale (provided as online supplemental material). A small number (< 10%) of EPA observations within the Transition to Discipline and Foundations of Discipline stages were rated "I had to do" (1 of 5) or "I had to talk them through" (2 of 5). Most (> 60%) of the EPAs observed at this stage were rated as "I had to be there just in case" (4 of 5) or "I didn't need to be there" (5 of 5). Substantially less data were available for the Core of Discipline stage, but the pattern was similar.

Correlation Data

The number of EM and pediatric EM rotations within each program demonstrated a strong correlation ($r = 0.83$, $P < .001$), with the average number of EPAs observed per resident.

Discussion

This article describes the first Canadian dataset representative of the national CBME rollout in any RCPSC specialty. Key findings include a substantial variability in the number of EPA observations and promotion timelines across programs, the promotion of most residents prior to achieving the recommended number of EPA observations, few ratings at the low end of the entrustment scale, and a strong correlation between the average number of EPA observations per resident and time spent on EM rotations.

Our findings may inform individual program improvement and the modification of our national assessment framework. For example, local implementation leaders with lower-than-expected EPA observations may identify ways to increase observation frequency by seeking advice from other programs. Simultaneously, programs may identify practical obstacles that will inform modifications of national standards. Overall, the frequency with which individual EPAs are assessed will have important implications for the operational aspects of this new assessment program.

The variability that we have identified highlights the possibility that trainee experience is highly heterogeneous. There could be numerous explanations for this (eg, varying levels of engagement, differences in teaching skillsets, amount of faculty development, etc), but compared to the previous time-based model where this variability was largely an undocumented problem, this new system allows us to

quantify trainee experiences and work toward greater standardization across programs.[3,26] Because this article outlines a single year of data from a single specialty, it is a starting point from which to evolve the assessment program, rather than an inculcation of the lack of fidelity of implementation with CBME in general.

Our data collection approach was different than those described elsewhere,[2–4,10,11] due to limitations in our ability to access the assessment data and the engagement of members of each programs' leadership in the research. Direct involvement of these key stakeholders in this process is likely to have focused our analysis on program-level metrics that are of relevance to them[26–28] and increased buy-in in the program evaluation process.[27–29] This will increase the likelihood that the results will be used by stakeholders as intended—to support the improvement of the participating programs.[20,22]

Our findings are also unique in that they incorporate unprocessed program-level assessment data (ie, EPA observation numbers and scores) and trainee progression data (ie, when trainees were promoted between levels). Previous literature from the ACGME utilized national data that was amalgamated from the reports of individual clinical competency committees after they had determined achievement for trainees.[2–4] As demonstrated recently in a subset of EM programs in the United States, there are discrepancies between reported data regarding trainee promotion[2] and the data acquired for local decision making.[10,11] This may suggest that human judgement allows for better representation of performance, adjusting for local culture and nuances. We feel that by monitoring both sets of data in tandem, broader questions about idiosyncratic or systemic biases could be elucidated.

The collection of unprocessed data also demonstrated a substantial amount of program-level variation. While some variability in EPA numbers is expected given local contexts, a 2-fold difference in the number of EPAs observed suggests substantial heterogeneity. This may be due to local engagement with CBME, or other factors may be at play as well (eg, in our analysis on the number of EM rotations in the first year was a key factor). Additional variability may have also resulted from piloting the assessment program, previous use of a workplace-based assessment program (3 sites), an earlier rollout date of the assessment program (2 sites), and technical difficulties with various learning management systems (reported by several programs). The use of a modified 5-point entrustment score at the University of Toronto (provided as online supplemental material) may have impacted EPA observation metrics from that site.

Similar to the work of Conforti and colleagues,[4] these early analyses may inform our specialty committee's evolution of our assessment program (eg, modify the EPA observation suggestions). However, with the additional context provided by seeing other programs' data and structural elements, this report may also inform local program-level reflections and changes to explore what program facets have positive or negative effects on EPA observations. For instance, data sharing and comparisons may help to identify successful local innovations that can be scaled nationally.

Our results raise additional questions. For example, there was a substantial delay in the promotion for many residents. While variability in promotion timelines is a feature of CBME,[15,30] the observed degree of variability suggests that either the assessment program is identifying residents who are falling behind early, or, perhaps more likely, variability in competence committee practices or promotion standards are impacting the rate of resident progress at this early stage. Promotions occurred more often in September, December, March, and June, suggesting that the timing of competence committee meetings may have impacted resident promotion timelines. Notably, very few EPAs were scored at low levels of the entrustment scale. This could be due to leniency or range restriction by assessors,[31] resident "gaming" of assessments to avoid low scores,[32,33] excellent preparation of learners by undergraduate medical training programs, or the assessment culture.[34,35]

Limitations

Our study contained only the initial quantitative data for the first year of our implementation. Moreover, manual data extraction can be error prone despite the efforts taken to ensure that it was checked locally prior to compilation. We also anticipate that our relatively small sample size, advances in faculty development,[36] and increasing comfort with the program of assessment may reduce the generalizability of our results over time. Another issue surrounded learning management systems: due to computer database interface issues 2 programs recorded ultrasound EPAs (Core of Discipline EPA 14), which were inaccessible to us at the time of this analysis. Inclusion of these items would have slightly increased the total number of core EPAs and EPAs per resident observed in these programs. Finally, 2 programs declined to participate—one due to philosophical differences surrounding data governance and another due to a transition in leadership (ie, no site lead was available to participate at the time of data collection). The total number of trainees within this group of non-

participatory programs was low (n = 9, or 11.5% of the total number of trainees nationally), and we believe it is unlikely that it would change our analyses.

Next Steps

The collection and analysis of program- and national-level assessment data is an important first step in evaluating the impact of our assessment program on training. While the investigation of higher-order outcomes in the educational (eg, pursuit of fellowships, etc) and clinical (eg, clinical competence, attending practice metrics, etc) realms has been proposed,[14] substantive variation in the fidelity of the implementation of CBME programs may make it difficult to attribute outcome differences to the assessment program.[7,12] The defining of educationally important and measurable outcomes will be critical for establishing a robust plan for evaluating CBME systems and has been initiated in parallel to this work.[14]

Moving forward, we hope to analyze person-level and narrative data. Person-level data could allow the evaluation of systemic biases (eg, race or gender bias) in the assessment data, determine the number of promotion data points that competency committees use to promote trainees, or evaluate the effects of curricular differences on EPA observations. The narrative data generated from a national assessment system may offer additional insights.[37–40] We anticipate that other specialties may utilize our data amalgamation methods to evaluate their own CBME assessment programs. Beyond program evaluation, the collected dataset could have significant research value, especially if linked to other datasets (eg, medical school training records, clinical outcome databases).[41,42]

Conclusions

In efforts to improve both program and national-level CBME assessment programs, we have shown that involving and engaging program-level educational leaders to collect and aggregate data can yield unique analytics that are useful to both local and national stakeholders and leaders. The findings in our evaluation study represent a new approach to integrating national and local program data to allow for improvement processes at both levels.

References

1. Frank JR, Snell L, Englander R, Holmboe ES. Implementing competency-based medical education: moving forward. *Med Teach*. 2017;39(6):568–573. doi:10.1080/0142159X.2017.1315069.
2. Santen SA, Yamazaki K, Holmboe ES, Yarris LM, Hamstra SJ. Comparison of male and female resident milestone assessments during emergency medicine residency training. *Acad Med*. 2020;95(2):263–268. doi: 10.1097/ACM.0000000000002988.
3. Hamstra SJ, Yamazaki K, Barton MA, Santen SA, Beeson MS, Holmboe ES. A national study of longitudinal consistency in ACGME milestone ratings by Clinical Competency Committees. *Acad Med*. 2019;94(10):1522–1531. doi:10.1097/acm.0000000000002820.
4. Conforti LN, Yaghmour NA, Hamstra SJ, Holmboe ES, Kennedy B, Liu JJ, et al. The effect and use of milestones in the assessment of neurological surgery residents and residency programs. *J Surg Educ*. 2018;75(1):147–155. doi:10.1016/j.jsurg.2017.06.001.
5. Chan T, Sherbino J, Collaborators M. The McMaster Modular Assessment Program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program. 2015;90(7):900–905. doi:10.1097/acm.0000000000000707.
6. Li S, Sherbino J, Chan TM. McMaster Modular Assessment Program (McMAP) through the years: residents' experience with an evolving feedback culture over a 3-year period. *AEM Educ Train*. 2017;1(1):5–14. doi:10.1002/AET2.10009.
7. Hall A, Rich J, Dagnone J, Weersink K, Caudle J, Sherbino J, et al. It's a marathon, not a sprint: rapid evaluation of CBME program implementation. *Acad Med*. 2020;95(5):786–793. doi:10.1097/ACM.0000000000003040.
8. Ross S, Binczyk NM, Hamza DM, Schipper S, Humphries P, Nichols D, et al. Association of a competency-based assessment system with identification of and support for medical residents in difficulty. *JAMA Netw Open*. 2018;1(7):e184581. doi:10.1001/jamanetworkopen.2018.4581.
9. Binczyk NM, Babenko O, Schipper S, Ross S. Unexpected result of competency-based medical education: 9-year application trends to enhanced skills programs by family medicine residents at a single institution in Canada. *Educ Prim Care*. 2019;30(3):1–6. doi:10.1080/14739879.2019.1573108.
10. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med*. 2017;177(5):651–657. doi:10.1001/jamainternmed.2016.9616.
11. Mueller AS, Jenkins T, Osborne M, Dayal A, O'Connor DM, Arora VM. Gender differences in attending physicians' feedback for residents in an emergency

medical residency program: a qualitative analysis. *J Grad Med Educ*. 2017;9(5):577–585. doi:10.4300/JGME-D-17-00126.1.

12. Mowbray CT, Holter MC, Teague GB, Bybee D. Fidelity criteria: development, measurement, and validation. *Am J Eval*. 2003;24(3):315–340. doi:10.1016/S1098-2140(03)00057-2.
13. Century J, Rudnick M, Freeman C. A framework for measuring fidelity of implementation: a foundation for shared language and accumulation of knowledge. *Am J Eval*. 2010;31(2):199–218. doi:10.1177/1098214010366173.
14. Chan TM, Paterson QS, Hall AK, Zaver F, Woods RA, Hamstra SJ, et al. Outcomes in the age of competency-based medical education: recommendations for emergency medicine training outcomes in the age of competency-based medical education. *CJEM*. 2020;22(2):204–214. doi:10.1017/cem.2019.491.
15. Sherbino J, Bandiera G, Doyle K, Frank JR, Holroyd BR, Jones G, et al. The competency-based medical education evolution of Canadian emergency medicine specialist training. *CJEM*. 2020;22(1):95–102. doi:10.1017/cem.2019.417.
16. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med*. 2012;87(10):1401–1407. doi:10.1097/ACM.0b013e3182677805.
17. MacEwan MJ, Dudek NL, Wood TJ, Gofton WT. Continued validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): use in the simulated environment. *Teach Learn Med*. 2016;28(1):72–79. doi:10.1080/10401334.2015.1107483.
18. Royal College of Physicians and Surgeons of Canada. Entrustable Professional Activity Guide: Emergency Medicine. https://cloudfront.ualberta.ca/-/media/medicine/departments/emergency-medicine/documents/epa-guide-emergency-med-e.pdf. Accessed June 3, 2020.
19. Hahn EJ, Noland MP, Rayens MK, Christie DM. Efficacy of training and fidelity of implementation of the life skills training program. *J Sch Health*. 2002;72(7):282–287. doi:10.1111/j.1746-1561.2002.tb01333.x.
20. Royal College of Physicians and Surgeons of Canada. Frank JR, Snell S, Sherbino J, eds. CanMEDS 2015 Physician Competency Framework. http://www.royalcollege.ca/rcsite/documents/canmeds/canmeds-full-framework-e.pdf. Accessed June 3, 2020.
21. Hauer KE, Chesluk B, Iobst W, Holmboe E, Baron RB, Boscardin CK, et al. Reviewing residents' competence: a qualitative study of the role of clinical competency committees in performance assessment. *Acad Med*. 2015;90(8):1084–1092. doi:10.1097/ACM.0000000000000736.
22. Kinnear B, Warm EJ, Hauer KE. Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. *Med Teach*. 2018;40(11):1110–1115. doi:10.1080/0142159X.2018.1474191.
23. Hauer KE, Cate O Ten, Boscardin CK, Iobst W, Holmboe E, et al. Ensuring resident competence: a narrative review of the literature on group decision-making to inform the work of Clinical Competency Committees. *J Grad Med Educ*. 2016;8(2):156–164. doi:10.4300/JGME-D-15-00144.1.
24. Chahine S, Cristancho S, Padgett J, Lingard L. How do small groups make decisions?: A theoretical framework to inform the implementation and study of clinical competency committees. *Perspect Med Educ*. 2017;6(3):192–198. doi:10.1007/s40037-017-0357-x.
25. Government of Canada. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans—TCPS 2 (2018). https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html. Accessed July 27, 2020.
26. Chen HT. The bottom-up approach to integrative validity: a new perspective for program evaluation. *Eval Program Plann*. 2010;33(3):205–214. doi:10.1016/j.evalprogplan.2009.10.002.
27. Azzam T. Evaluator responsiveness to stakeholders. *Am J Eval*. 2010;31(1):45–65. doi:10.1177/1098214009354917.
28. Israel BA, Cummings KM, Dignan MB, Heaney CA, Perales DP, Simons-Morton BG, et al. Evaluation of health education programs: current assessment and future directions. *Heal Educ Q*. 1995;22(3):364–389. doi:10.1177/109019819402200308.
29. Franz NK. The data party: involving stakeholders in meaningful data analysis. *J Extension*. 2013;51(1):1–2.
30. ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ*. 2005;39(12):1176–1177. doi:10.1111/j.1365-2929.2005.02341.x.
31. Govaerts MJB, Schuwirth LWT, van der Vleuten CPM, Muijtjens AM. Workplace-based assessment: effects of rater expertise. *Adv Heal Sci Educ*. 2011;16(2):151–165. doi:10.1007/s10459-010-9250-7.
32. 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(9):953–962. doi:10.1111/medu.13380.
33. Acai A, Li SA, Sherbino J, Chan TM. Attending emergency physicians' perceptions of a programmatic workplace-based assessment system: The McMaster Modular Assessment Program (McMAP). *Teach Learn*

Med. 2019;31(4):434–444. doi:10.1080/10401334.2019.1574581.
34. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med*. 2005;80(10 Suppl):84–87. doi:10.1097/00001888-200510001-00023.
35. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76–85. doi:10.1111/medu.13645.
36. Stefan A, Hall JN, Sherbino J, Chan TM. Faculty development in the age of competency-based medical education: a needs assessment of Canadian emergency medicine faculty and senior trainees. *CJEM*. 2019;21(4):527–534. doi:10.1017/cem.2019.343.
37. 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.
38. Ginsburg S, van der Vleuten CPM, Eva KW. The hidden value of narrative comments for assessment: a quantitative reliability analysis of qualitative data. *Acad Med*. 2017;92(11):1617–1621. doi:10.1097/ACM.0000000000001669.
39. 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.
40. Ginsburg S, van der Vleuten C, Eva KW, Lingard L. Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. *Adv Heal Sci Educ Theory Pract*. 2016;21(1):175–188. doi:10.1007/s10459-015-9622-0.
41. Ellaway RH, Topps D, Pusic M. Data, big and small: emerging challenges to medical education scholarship. *Acad Med*. 2019;94(1):31–36. doi:10.1097/ACM.0000000000002465.
42. Ellaway RH, Pusic MV, Galbraith RM, Cameron T. Developing the role of big data and analytics in health professional education. *Med Teach*. 2014;36(3):216–222. doi:10.3109/0142159X.2014.874553.

All authors are collaborators from the Canadian Data Research for Evaluation and Analytics in Medicine (CanDREAM) team. **Brent Thoma, MD, MA, MSc, FRCPC, DRCPSC,** is Associate Professor, Department of Emergency Medicine, University of Saskatchewan, Saskatoon, SK, Canada, and Clinician Educator, Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada; **Andrew K. Hall, MD, FRCPC, MMEd,** is Associate Professor, Department of Emergency Medicine, Queen's University, Kingston, Ontario, Canada, and Clinician Educator, Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada; **Kevin Clark, MD, MMEd, FRCPC,** is Program Director, University of British Columbia Emergency Medicine Residency Program, and Clinical Associate Professor, University of British Columbia, Vancouver, British Columbia, Canada; **Nazanin Meshkat MD, FRCPC, MHSc,** is Associate Professor, Division of Emergency Medicine, University of Toronto, Toronto, Ontario, Canada, and Program Director and CBD Lead, Royal College Emergency Medicine Residency Program, Ottawa, Ontario, Canada; **Warren J. Cheung, MD, MMEd, FRCPC,** is Assistant Professor, Director of Assessment, and Associate Director of Education Innovation, Department of Emergency Medicine, University of Ottawa, Ottawa, Ontario, Canada, and CanMEDS Clinician Educator, Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada; **Pierre Desaulniers, MD, CSPQ,** is Associate Professor of Clinical Medicine and Director of the Emergency Medicine Residency Program, and Chairman, Patient Safety and Resident Safety Committee, University of Montreal, Montreal, Canada; **Cheryl ffrench, MD, FRCPC,** is Assistant Professor, Department of Emergency Medicine, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, and Program Director and *Competency by Design* Lead, Emergency Medicine Residency Training Program, University of Manitoba Royal College, Winnipeg, Manitoba, Canada; **Allison Meiwald, BN, MD, MEd, FRCPC,** is Assistant Professor, Division of Emergency Medicine, Department of Medicine, Schulich School of Medicine and Dentistry, and Assistant Program Director, Emergency Medicine Residency Program, Western University, London, Ontario, Canada; **Christine Meyers, MD, FRCPC,** is Faculty Lecturer, Department of Emergency Medicine, McGill University and Residency Education Coordinator for Emergency Medicine, Royal Victoria Hospital, Barrie, Ontario, Canada; **Catherine Patocka, MDCM, MHPE, FRCPC,** is Clinical Assistant Professor, Department of Emergency Medicine, Cumming School of Medicine, University of Calgary and Program Director, and the Division of Education and Innovation, Emergency Medicine Residency Training Program, University of Calgary and the Division of Education and Innovation, Calgary, Alberta, Canada; **Lorri Beatty, MD, FRCPC,** is Assistant Professor, Department of Emergency Medicine, and Assistant Program Director, Emergency Medicine Residency Training Program, Dalhousie University, Halifax, Nova Scotia, Canada; and **Teresa M. Chan, MD, MHPE, FRCPC, DRCPSC,** is Associate Professor, Divisions of Emergency Medicine and Education and Innovation, Department of Medicine, Assistant Dean of Faculty Development, McMaster University Faculty of Health Sciences, Program Director, Clinician Educator Area of Focused Competency Program, McMaster University, Teacher, Michael G. DeGroote School of Medicine, Adjunct Scientist, McMaster Education Research, Innovation, and Theory (MERIT) Center, Specialty Committee Chair, Clinician Educator Area of Focused Competency, Royal College of Physicians and Surgeons of Canada, and Associate Editor, *Journal of Graduate Medical Education*.

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 as part of a podium presentation at the International Conference of Residency Education, Ottawa, Canada, September 26–28, 2019.

Received November 12, 2019; revisions received February 11, 2020, and May 17, 2020; accepted May 20, 2020.