

Using Learning Analytics to Examine Differences in Assessment Forms From Continuous Versus Episodic Supervisors of Family Medicine Residents

Ann S. O. Lee¹, MD, MEd

Christopher Donoff, MSc

Shelley Ross², PhD

ABSTRACT

Background It is assumed that there is a need for continuity of supervision within competency-based medical education, despite most evidence coming from the undergraduate medical education rather than the graduate medical education (GME) context. This evidence gap must be addressed to justify the time and effort needed to redesign GME programs to support continuity of supervision.

Objective To examine differences in assessment behaviors of continuous supervisors (CS) versus episodic supervisors (ES), using completed formative assessment forms, FieldNotes, as a proxy.

Methods The FieldNotes CS- and ES-entered for family medicine residents (N=186) across 3 outpatient teaching sites over 3 academic years (2015-2016, 2016-2017, 2017-2018) were examined using 2-sample proportion z-tests to determine differences on 3 FieldNote elements: competency (Sentinel Habit [SH]), Clinical Domain (CD), and Progress Level (PL).

Results Sixty-nine percent (6104 of 8909) of total FieldNotes were analyzed. Higher proportions of CS-entered FieldNotes indicated SH3 (Managing patients with best practices), $z=-3.631$, $P<.0001$; CD2 (Care of adults), $z=-8.659$, $P<.0001$; CD3 (Care of the elderly), $z=-4.592$, $P<.0001$; and PL3 (Carry on, got it), $z=-4.482$, $P<.0001$. Higher proportions of ES-entered FieldNotes indicated SH7 (Communication skills), $z=4.268$, $P<.0001$; SH8 (Helping others learn), $z=20.136$, $P<.0001$; CD1 (Doctor-patient relationship/ethics), $z=14.888$, $P<.0001$; CD9 (Not applicable), $z=7.180$, $P<.0001$; and PL2 (In progress), $z=5.117$, $P<.0001$.

Conclusions The type of supervisory relationship impacts assessment: there is variability in which competencies are paid attention to, which contexts or populations are included, and which progress levels are chosen.

Introduction

Continuity of supervision is one of the assumptions of competency-based medical education (CBME),¹⁻⁴ but is the call for continuity of supervision supported by evidence? Evidence supporting the benefits of continuity of supervision for competency assessments comes primarily from undergraduate medical education (UME),⁵⁻¹² with little evidence from graduate medical education (GME).¹³ Thus, it is not clear whether the investment of resources and time to restructure GME to enhance continuity of supervision is justified.

The benefits of integrating continuity of supervision into CBME assessments derives from primarily UME work demonstrating relationships between supervisors and learners can support learners' attitudes toward assessment,^{10,14,15} willingness to seek out feedback,¹⁶ and incorporation of feedback into learning.¹⁷ Relationships have also been highlighted for effective coaching,^{18,19} and supervisor-learner

relationships appear integral to the development of trust.²⁰ There is also heterogeneity in the definitions and contexts for continuity of supervision in studies of competency assessments. UME studies focus on yearlong longitudinal integrated clerkships versus clerkship rotations.²¹ In GME, even 5 sessions with the same supervisor may be considered "high continuity."²² Without understanding different supervisory relationships and their effects on workplace-based assessments in GME, conclusions from feedback and assessment in CBME may be flawed.

This study used learning analytics, the collection and analysis of education data, to compare existing workplace-based resident assessments for characteristics of assessments performed in continuous versus episodic supervisory relationships.^{23,24}

Methods

Setting and Participants

This study was conducted in 2020 in a large university-based Canadian 2-year family medicine residency program (note: 2 years is standard in Canada), with 70 to 80 residents per year (60 to 65 in the urban stream and 12 to 16 in the rural stream).

DOI: <http://dx.doi.org/10.4300/JGME-D-21-00832.1>

Editor's Note: The online version of this article contains further data from the study.

Upon entry, residents are assigned to a continuity supervisor for the 2 years. During the first year, residents are assigned to a 6-month continuous experience with the continuity supervisor and then a scheduled half-day per week in this “home site,” with the same supervisor for the remainder of residency. All other supervisors are non-continuous, “episodic” supervisors, including other family medicine supervisors.

Data Source

The residency program uses electronic forms called FieldNotes for formative, work-based assessments.^{25,26} Supervisors complete a FieldNote after direct observation of a resident by selecting one each of the following elements: competency (labeled Sentinel Habits [SH]), which is based on the Assessment Objectives for Certification in Family Medicine (provided as online supplementary data)²⁷; clinical context or population (labeled Clinical Domain [CD] of family medicine; TABLE); and judgement of the level of competence demonstrated (Progress Level [PL]; TABLE). We looked at SH, CD,

Objectives

Examining differences in assessment behaviors of continuous supervisors versus episodic supervisors, using completed formative assessment forms, FieldNotes, to help determine whether time and effort needed to redesign graduate medical education (GME) programs to support continuity of supervision can be justified.

Findings

At the GME level, type of supervisory relationship impacts assessment in terms of competencies paid attention to, contexts or populations included, and progress levels chosen.

Limitations

Our study was conducted at a single institution and specialty, which may limit generalizing particularly to programs with longer training programs where there is more opportunity to interact with residents over time.

Bottom Line

This study provides evidence supporting the benefits of continuity of supervision for competency assessments. Without understanding different supervisory relationships and their effects on workplace-based assessments in GME, conclusions from feedback and assessment in competency-based medical education may be flawed.

TABLE
Variables of Interest From FieldNotes as Defined in the Family Medicine Residency Program

No.	Sentinel Habit (Competency)
SH1	Incorporates the patient's experience and context into problem identification and management
SH2	Generates relevant hypotheses resulting in a safe and prioritized differential diagnosis
SH3	Manages patients using available best practices
SH4	Selects and attends to the appropriate focus and priority in a situation
SH5	Uses generic key features when performing a procedure (such as knowing indications and contraindications, etc)
SH6	Demonstrates respect and/or responsibility
SH7	Verbal or written communication is clear and timely
SH8	Helps others learn (teaches colleagues and patients in a useful manner)
SH9	Promotes effective practice quality (eg, EMR use, practice audits)
SH10	Seeks and responds appropriately to guidance and feedback
No.	Clinical Domain of Family Medicine
CD1	Doctor-patient relationship/ethics
CD2	Care of adults
CD3	Care of children and adolescents
CD4	Care of the elderly
CD5	Care of the vulnerable and underserved
CD6	Maternity/newborn care
CD7	Palliative/end of life care
CD8	Surgical and procedural skills
CD9	Not applicable
No.	Progress Level (Assessor Judgement of Competence Demonstrated)
PL1	Stop, important correction
PL2	In progress
PL3	Carry on, got it

Abbreviations: SH, Sentinel Habit; EMR, electronic medical record; CD, Clinical Domain; PL, Progress Level.

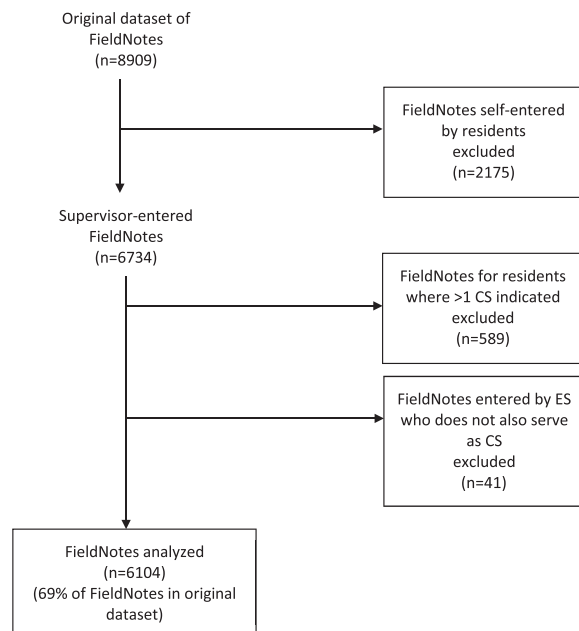


FIGURE 1
Flow Diagram of Inclusion and Exclusion of FieldNotes
Abbreviations: ES, episodic supervisor, CS, continuous supervisor.

and PL on all FieldNotes entered online for all residents (N=186) by their supervisors (N=79) across 3 academic teaching sites and 3 consecutive cohorts within academic years 2015-2016, 2016-2017, and 2017-2018. FieldNotes were de-identified by replacing resident and supervisor names with a unique identifier code. FieldNotes were also coded according to the type of supervision (continuous supervisor [CS] or episodic supervisor [ES]).

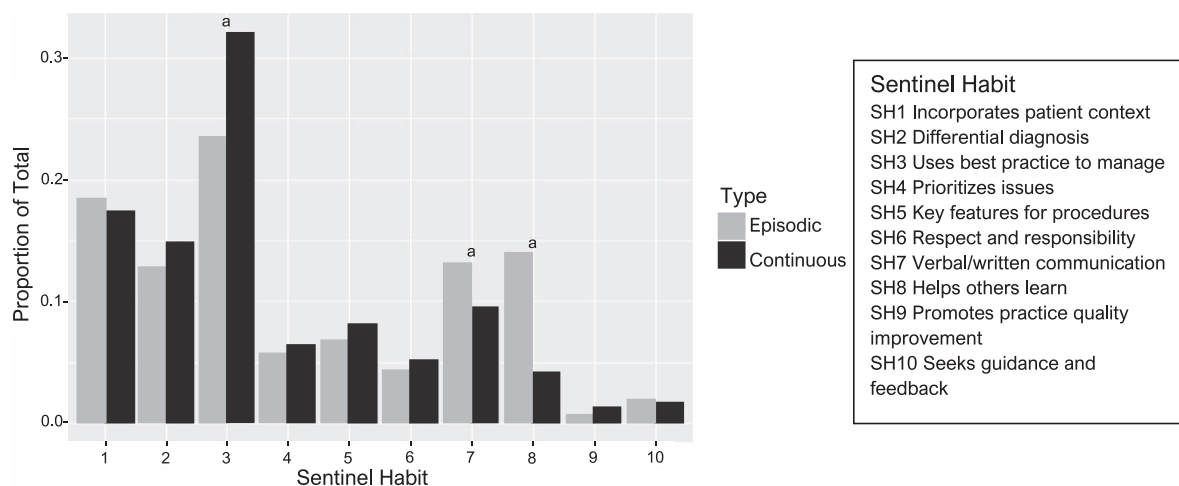


FIGURE 2
Distribution of Proportion of Sentinel Habits Between Continuous and Episodic Supervisor-Entered FieldNotes
^a $P < .0001$.

Statistical Analysis

Our initial exploratory data analysis consisted of visualizations and 2-sample proportion z-tests (using Bonferroni correction for Type I error, ie, $\alpha/n=10$, with $\alpha=.05$ and $n=10$, the number of comparisons we made) to compare frequencies of SHs, CDs, and PLs between FieldNotes completed by CS versus ES.²⁸ The Bonferroni correction provided a more rigorous correction than the Tukey test which tolerates Type I errors, but more generous than the conservative Scheffé's method, providing a balanced control for Type I errors. A test of multicollinearity on the variables SH, CD, and PL using the variance inflation factor was performed. None of the variance inflation factor values were large enough to warrant a concern that these variables are highly correlated with each other, supporting our assumption of independence. All statistical analyses were conducted using RStudio.

This study was approved by the University of Alberta Health Research Ethics Board.

Results

The original dataset included 8909 FieldNotes. We excluded 2175 notes that were self-entered by a resident or by a supervisor who did not serve as a CS for any resident, and where there was uncertainty about which supervisor was assigned to a resident. The analyses included 6104 FieldNotes (69% of total FieldNotes; FIGURE 1).

Analysis of 6104 FieldNotes showed differences in the proportion of SHs, CDs, and PLs between FieldNotes entered by CS and ES. Using the 2-sample test of proportions, higher proportions of CS-entered

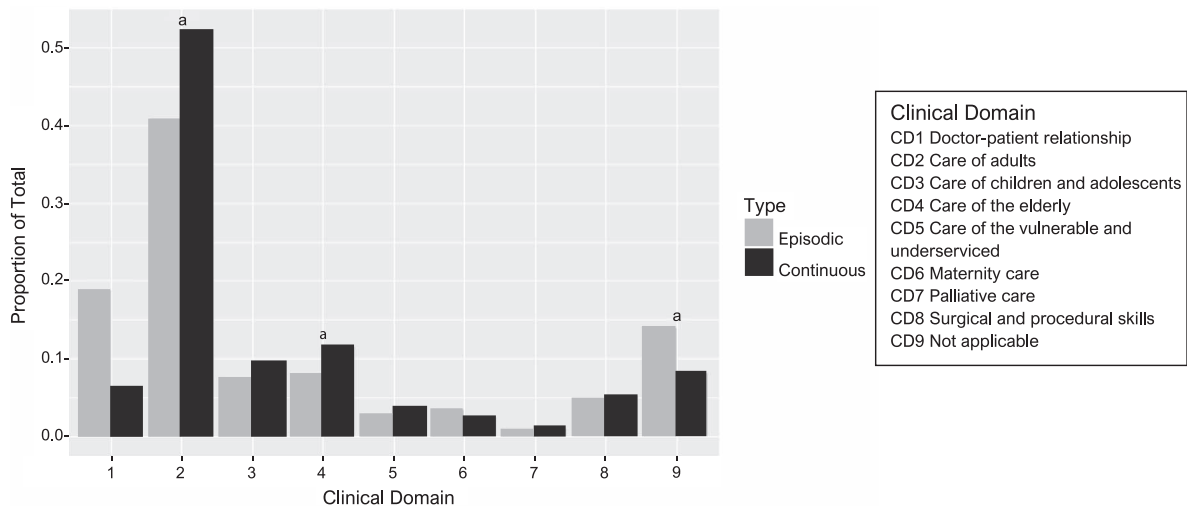


FIGURE 3
Distribution of Proportion of Clinical Domains Between Continuous and Episodic Supervisor-Entered FieldNotes
^a $P < .0001$.

FNs indicated SH3 (Managing patients with best practices), $z = -3.631$, $P < .0001$; CD2 (Care of adults), $z = -8.659$, $P < .0001$; CD3 (Care of the elderly), $z = -4.592$, $P < .0001$; and PL3 (Carry on, got it), $z = -4.482$, $P < .0001$ (FIGURES 2-4). Higher proportions of ES-entered FieldNotes indicated SH7 (Communication skills), $z = 4.268$, $P < .0001$; SH8 (Helping others learn), $z = 20.136$, $P < .0001$; CD1 (Doctor-patient relationship/ethics), $z = 14.888$, $P < .0001$; CD9 (Not applicable), $z = 7.180$, $P < .0001$; and PL2 (In progress), $z = 5.117$, $P < .0001$.

Discussion

Although both continuous and episodic supervisors completed FieldNotes across the range of Sentinel Habits, Clinical Domains, and Progress Levels, we found differences in specific competencies, clinical context or population, and competence levels assessed by CS versus ES.

Assessment information collected with FieldNotes is used in a similar way to the information collected by entrustable professional activities (EPAs) assessments in many CBME programs.^{26,29} With both

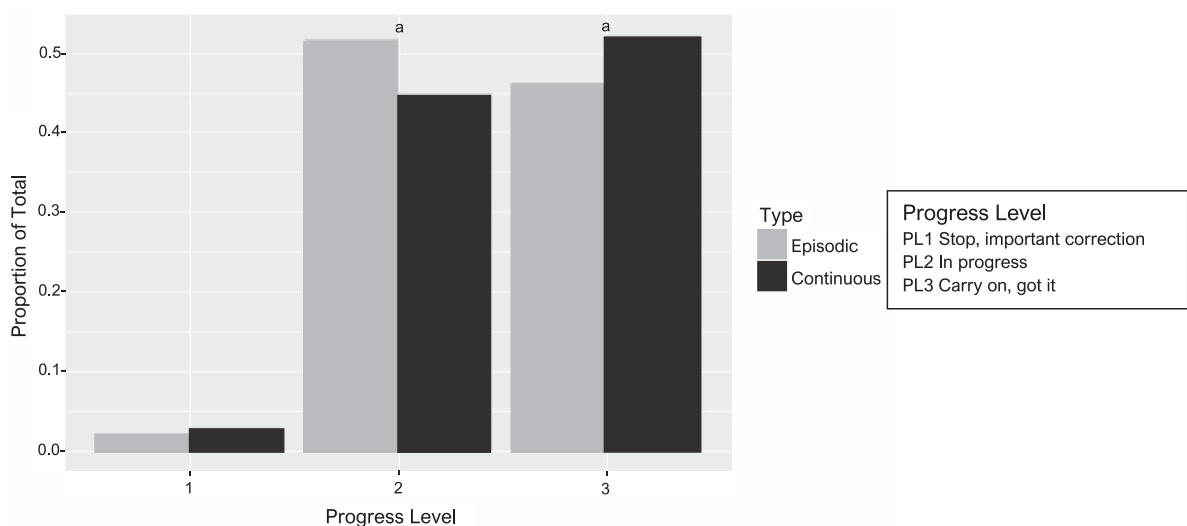


FIGURE 4
Distribution of Proportion of Progress Levels Between Continuous and Episodic Supervisor-Entered FieldNotes
^a $P < .0001$.

FieldNotes and EPAs, residents are assessed by multiple assessors over multiple instances, and these assessments lead to summative decisions. The study results suggest that different supervisory relationships, by focusing on different resident behaviors, may enhance the assessment program. The differences found between CS and ES suggest that supervisory relationship may affect what assessors choose to assess. Thus here, for ES relationships, assessors were more likely to assess SHs where competency can be demonstrated in one or a few observations, such as Verbal or written communications (SH7) or Doctor-patient relationships/ethics (CD1).

In contrast, CS-entered FieldNotes were more likely to assess managing patients using available best practices (SH3), the application of medical knowledge in the management of patients. This is a complex competency that incorporates multiple elements and likely requires that a resident be observed repeatedly over time across multiple contexts. This is more likely to occur when there is a continuous supervisory relationship. Further, the higher proportion of FieldNotes from CS with the highest judgement of competence (PL3: Carry on, got it) likely reflects trust that supervisors develop when they have a sustained, longitudinal relationship with a resident.³⁰

The difference in judgement of competence is particularly important as it suggests supervisors are less likely to indicate that a resident has demonstrated competence when they only work with a resident a few times or intermittently. This finding is important to explore further. Currently, many GME programs incorporate primarily ES because of curricular and workplace-based structures.³¹⁻³³ If our finding is found to be common across programs, residents may be less likely to be deemed competent or entrusted because they are being assessed by an ES who has not had the opportunity to develop a relationship that allows for understanding of a resident's progression toward competence and for the development of trust over multiple entrustment experiences. Assessment of competence in GME can be high stakes, even leading to shortening or lengthening of training. It is important that we have a clear understanding of how different supervisory relationships impact assessment behavior.

This study is limited by use of a single institution and specialty, which may limit generalizing particularly to programs with longer training programs. We looked for patterns of differences in assessment between CS and ES and did not control for other factors, such as the type of clinical experience, assessor training or years in program, and other

factors. In addition to workplace-based assessment, the FieldNotes included assessments of simulated patient care sessions and resident teaching sessions, which likely present fewer domains for assessment. Also, this study compared the frequency of assessment areas, not the quality of supervisory assessments. Future research will use learning analytics to further examine similarities and differences between CS and ES assessments, including amount and quality of feedback captured on the FieldNotes, and whether there is evidence that feedback loops are closed.³⁴ We will also collaborate with other residency programs who have similar datasets to look at whether our findings can be replicated using their datasets.

Conclusions

The type of supervisory relationship impacts assessment: for example, supervisors who had a continuous relationship with a learner were more likely to comment on complex competencies, while episodic supervisors were more likely to comment on competencies that require fewer observations. Depending on the type of supervisory relationship, there is also variability in the contexts or populations included and in the progress levels chosen on FieldNotes.

References

1. Holmboe ES, Ward DS, Reznick RK, et al. Faculty development in assessment: the missing link in competency-based medical education. *Acad Med.* 2011;86(4):460-467. doi:10.1097/ACM.0b013e31820cb2a7
2. Hawkins RE, Welcher CM, Holmboe ES, et al. Implementation of competency-based medical education: are we addressing the concerns and challenges? *Med Educ.* 2015;49(11):1086-1102. doi:10.1111/medu.12831
3. 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
4. Englander R, Carraccio C. A lack of continuity in education, training, and practice violates the "do no harm" principle. *Acad Med.* 2018;93(suppl 3):12-16. doi:10.1097/ACM.0000000000002071
5. Ogur B, Hirsh D, Krupat E, Bor D. The Harvard Medical School-Cambridge integrated clerkship: an innovative model of clinical education. *Acad Med.* 2007;82(4):397-404. doi:10.1097/ACM.0b013e31803338f0

6. Bell SK, Krupat E, Fazio SB, Roberts DH, Schwartzstein RM. Longitudinal pedagogy: a successful response to the fragmentation of the third-year medical student clerkship experience. *Acad Med*. 2008;83(5):467-475. doi:10.1097/ACM.0b013e31816bdad5
7. Hauer KE, Mazotti L, O'Brien B, Hemmer PA, Tong L. Faculty verbal evaluations reveal strategies used to promote medical student performance. *Med Educ Online*. 2011;16. doi:10.3402/meo.v16i0.6354
8. Hauer KE, O'Brien BC, Hansen LA, et al. More is better: students describe successful and unsuccessful experiences with teachers differently in brief and longitudinal relationships. *Acad Med*. 2012;87(10):1389-1396. doi:10.1097/ACM.0b013e31826743c3
9. Walters L, Greenhill J, Richards J, et al. Outcomes of longitudinal integrated clinical placements for students, clinicians and society. *Med Educ*. 2012;46(11):1028-1041. doi:10.1111/j.1365-2923.2012.04331.x
10. Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. *Med Educ*. 2013;47(4):362-374. doi:10.1111/medu.12087
11. Thistlethwaite JE, Bartle E, Chong AAL, et al. A review of longitudinal community and hospital placements in medical education: BEME Guide No. 26. *Med Teach*. 2013;35(8):e1340-e1364. doi:10.3109/0142159X.2013.806981
12. Cuncic C, Regehr G, Frost H, Bates J. It's all about relationships: a qualitative study of family physicians' teaching experiences in rural longitudinal clerkships. *Perspect Med Educ*. 2018;7(2):100-109. doi:10.1007/s40037-018-0416-y
13. Lee AS, Ross S. Continuity of supervision: does it mean what we think it means? *Med Educ*. 2021;55(4):448-454. doi:10.1111/medu.14378
14. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ*. 2019;53(1):76-85. doi:10.1111/medu.13645
15. Dijksterhuis MGK, Schuwirth LWT, Braat DDM, Teunissen PW, Scheele F. A qualitative study on trainees' and supervisors' perceptions of assessment for learning in postgraduate medical education. *Med Teach*. 2013;35(8):e1396-e1402. doi:10.3109/0142159X.2012.756576
16. Playford D, Kirke A, Maley M, Worthington R. Longitudinal assessment in an undergraduate longitudinal integrated clerkship: the mini Clinical Evaluation Exercise (mCEx) profile. *Med Teach*. 2013;35(8):e1416-e1421. doi:10.3109/0142159X.2013.778392
17. Schut S, Driessen E, van Tartwijk J, van der Vleuten C, Heeneman S. Stakes in the eye of the beholder: an international study of learners' perceptions within programmatic assessment. *Med Educ*. 2018;52(6):654-663. doi:10.1111/medu.13532
18. Voyer S, Cuncic C, Butler DL, MacNeil K, Watling C, Hatala R. Investigating conditions for meaningful feedback in the context of an evidence-based feedback programme. *Med Educ*. 2016;50(9):943-954. doi:10.1111/medu.13067
19. Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Med Educ*. 2014;48(7):713-723. doi:10.1111/medu.12407
20. Hauer KE, Oza SK, Kogan JR, et al. How clinical supervisors develop trust in their trainees: a qualitative study. *Med Educ*. 2015;49(8):783-795. doi:10.1111/medu.12745
21. Mazotti L, O'Brien B, Tong L, Hauer KE. Perceptions of evaluation in longitudinal versus traditional clerkships. *Med Educ*. 2011;45(5):464-470. doi:10.1111/j.1365-2923.2010.03904.x
22. Cheung WJ, Dudek NL, Wood TJ, Frank JR. Supervisor-trainee continuity and the quality of work-based assessments. *Med Educ*. 2017;51(12):1260-1268. doi:10.1111/medu.13415
23. Chan T, Sebok-Syer S, Thoma B, Wise A, Sherbino J, Pusic M. Learning analytics in medical education assessment: the past, the present, and the future. *AEM Educ Train*. 2018;2(2):178-187. doi:10.1002/aet2.10087
24. Thoma B, Warm E, Hamstra SJ, et al. Next steps in the implementation of learning analytics in medical education: consensus from an international cohort of medical educators. *J Grad Med Educ*. 2020;12(3):303-311. doi:10.4300/JGME-D-19-00493.1
25. Donoff MG. Field notes: assisting achievement and documenting competence. *Can Fam Physician Med Fam Can*. 2009;55(12):1260-1262, e100-e102.
26. Ross S, Poth CN, Donoff M, et al. Competency-based achievement system: using formative feedback to teach and assess family medicine residents' skills. *Can Fam Physician*. 2011;57(9):e323-e330.
27. Crichton T, Schultz K, Lawrence K, et al. Assessment Objectives for Certification in Family Medicine. Accessed May 21, 2022. <https://www.cfpc.ca/CFPC/media/Resources/Examinations/Assessment-Objectives-for-Certification-in-FM-full-document.pdf>
28. Velleman PF, Hoaglin DC. Exploratory Data Analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, eds. *APA Handbook of Research Methods in Psychology*. Vol 3. Washington, DC: American Psychological Association; 2012:51-70.
29. 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

30. Duijn CCMA, Welink LS, Bok HGJ, ten Cate O. When to trust our learners? Clinical teachers' perceptions of decision variables in the entrustment process. *Perspect Med Educ*. 2018;7(3):192-199. doi:10.1007/s40037-018-0430-0
31. Harris P, Bhanji F, Topps M, et al. Evolving concepts of assessment in a competency-based world. *Med Teach*. 2017;39(6):603-608. doi:10.1080/0142159X.2017.1315071
32. Bernabeo EC, Holtman MC, Ginsburg S, Rosenbaum JR, Holmboe ES. Lost in transition: the experience and impact of frequent changes in the inpatient learning environment. *Acad Med*. 2011;86(5):591-598. doi:10.1097/ACM.0b013e318212c2c9
33. ten Cate O, Hart D, Ankel F, et al. Entrustment decision making in clinical training. *Acad Med*. 2016;91(2):191-198. doi:10.1097/ACM.0000000000001044
34. Carless D. Feedback loops and the longer-term: towards feedback spirals. *Assess Eval High Educ*. 2019;44(5):705-714. doi:10.1080/02602938.2018.1531108



Ann S. O. Lee, MD, MEd, is Assistant Professor, Department of Family Medicine, University of Alberta, Edmonton, Alberta, Canada; **Christopher Donoff, MSc**, is Junior Data Scientist, Blackline Safety, Calgary, Alberta, Canada; and **Shelley Ross, PhD**, is Professor, Department of Family Medicine, University of Alberta Edmonton, Alberta, Canada.

Funding: This work was funded by the Social Sciences and Humanities Research Council of Canada.

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

This work was previously presented as a poster abstract at the virtual International Conference on Residency Education (ICRE), September 25, 2020, and as an education free standing paper at the virtual Family Medicine Forum, November 4, 2020.

The authors would like to thank Delane Linkiewicz for support in preparing data for analyses.

Corresponding author: Ann S. O. Lee, MD, MEd, University of Alberta, Edmonton, Alberta, Canada, ann.lee@ualberta.ca, Twitter @AnnLeeMedEd

Received August 24, 2021; revisions received January 29, 2022, and June 25, 2022; accepted June 28, 2022.