

Which Emergency Medicine Milestone Sub-competencies are Identified Through Narrative Assessments?

David Diller, MD*
Shannon Cooper, MD†
Aarti Jain, MD*
Chun Nok Lam, MPH*
Jeff Riddell, MD*

*LAC+USC Medical Center, Keck School of Medicine of the University of Southern California, Department of Emergency Medicine, Los Angeles, California
†Henry Ford Allegiance Health, Department of Emergency Medicine, Jackson, Michigan

Section Editor: Jonathan Fisher, MD, MPH

Submission history: Submitted July 14, 2019; Revision received December 3, 2019; Accepted December 4, 2019

Electronically published [date]

Full text available through open access at http://escholarship.org/uc/uciem_westjem

DOI: 10.5811/westjem.2019.12.44468

Introduction: Evaluators use assessment data to make judgments on resident performance within the Accreditation Council for Graduate Medical Education (ACGME) milestones framework. While workplace-based narrative assessments (WBNA) offer advantages to rating scales, validity evidence for their use in assessing the milestone sub-competencies is lacking. This study aimed to determine the frequency of sub-competencies assessed through WBNA in an emergency medicine (EM) residency program.

Methods: We performed a retrospective analysis of WBNA of postgraduate year (PGY) 2-4 residents. A shared mental model was established by reading and discussing the milestones framework, and we created a guide for coding WBNA to the milestone sub-competencies in an iterative process. Once inter-rater reliability was satisfactory, raters coded each WBNA to the 23 EM milestone sub-competencies.

Results: We analyzed 2517 WBNA. An average of 2.04 sub-competencies were assessed per WBNA. The sub-competencies most frequently identified were multitasking, medical knowledge, practice-based performance improvement, patient-centered communication, and team management. The sub-competencies least frequently identified were pharmacotherapy, airway management, anesthesia and acute pain management, goal-directed focused ultrasound, wound management, and vascular access. Overall, the frequency with which WBNA assessed individual sub-competencies was low, with 14 of the 23 sub-competencies being assessed in less than 5% of WBNA.

Conclusion: WBNA identify few milestone sub-competencies. Faculty assessed similar sub-competencies related to interpersonal and communication skills, practice-based learning and improvement, and medical knowledge, while neglecting sub-competencies related to patient care and procedural skills. These findings can help shape faculty development programs designed to improve assessments of specific workplace behaviors and provide more robust data for the summative assessment of residents. [West J Emerg Med. 2020;21(1):173-179.]

INTRODUCTION

In 2012, the Accreditation Council for Graduate Medical Education (ACGME) developed educational milestones to serve as the primary framework for competency-based assessment in graduate medical education.¹ These educational milestones were framed within specialty-specific sub-

competencies, with each sub-competency belonging to one of six previously established ACGME core competencies.² A central tenet to the milestones framework is the emphasis on resident trainee assessment based on observable performance and behaviors.³ While many workplace-based assessment strategies have been piloted, a comprehensive validated

approach to resident assessment within the milestones framework has yet to be developed.⁴⁻⁶

Workplace-based narrative assessment (WBNA), also known as the in-training evaluation report (ITER),⁷ uses descriptive commentary for performance assessment and has been proposed as an alternative method to checklists and rating scales.⁷⁻⁹ Through descriptive commentary, WBNA provides assessors with a version of assessment without the constraints of pre-selected ratings or options, theoretically allowing for a more robust analysis. WBNA can exist independently as an evaluation form, or in combination with checklists or rating scales as a hybrid evaluation model.

The benefits of WBNA in medical education are well documented. WBNA have been shown to be useful in ranking trainees,¹⁰ detecting learners who are experiencing difficulty,¹¹ identifying milestone sub-competencies that are more difficult to assess,¹⁰⁻¹² and predicting the need for resident remediation.¹³ In addition to providing a richer data source, narrative assessments are also appreciated by learners.^{14,15} Whether used in addition to anchor-based rating tools¹⁶ or as an independent assessment method, descriptive commentary can be a reliable method of assessment that influences faculty judgment on global resident performance.⁷

Despite these benefits, validity evidence for using WBNA as a method for assessing milestone sub-competencies in graduate medical education is lacking. The prevalence of vague comments, such as “hard worker” and “pleasant to work with,” are well-documented,¹⁷⁻²¹ and it is unclear how beneficial these comments are in assessing learners within the milestones framework. Furthermore, while contextual framing and faculty development can provide more robust narrative assessments in terms of both quantity and quality of comments,⁹ the frequency with which WBNA comment on specific milestone sub-competencies remains unknown. Without knowing which milestone sub-competencies are being assessed, program directors and clinical competency committees (CCC) may be left to assume competence in a broad range of skills, despite a lack of explicit evidence to support those conclusions.⁸

This study aims to determine the frequency of milestone sub-competencies assessed through semi-annual WBNA in an emergency medicine (EM) residency program.

METHODS

Settings and Participants

Faculty at the LAC+USC EM residency program complete semi-annual WBNA on residents with whom they have worked over the prior six-month period. An internally created online form through the education management platform MyEvaluations allows faculty to provide descriptive responses to two prompts: “Please describe at least one area of strength for this resident” and “Please describe at least one area for potential improvement for this resident.” Faculty WBNA are encouraged but not mandatory, and not all faculty complete

Educational Research Capsule Summary

What do we already know about this issue?

Narrative assessments are a commonly used evaluation tool for making judgments on resident clinical performance.

What was the research question?

What milestone sub-competencies are assessed through narrative assessments in an EM residency program?

What was the major finding of the study?

Unstructured narrative assessments identified relatively few milestone sub-competencies.

How does this improve population health?

These results can improve evaluation tool design and faculty development to improve the validity for narrative assessments within the Milestones framework.

WBNA on all residents. No formal training exists for faculty regarding milestone sub-competencies or workplace-based assessment strategies.

Study Design

We performed a retrospective analysis of the WBNA of postgraduate year (PGY) 2-4 residents completed between the second semester of 2016 and the first semester of 2017. WBNA of PGY-1 residents and second semester PGY-4 residents were excluded due to limited faculty contact with PGY-1 residents and a hypothesized concern from the study authors regarding a lack of critical assessment of PGY-4 residents during their final semester of training. The local institutional review board determined the study was exempt.

Protocol

Author DD collated, de-identified, and randomized the selected WBNA. Three authors (SC, AJ, JR), blinded to both the identity of the faculty assessor and resident being assessed, reviewed the WBNA to determine whether the comments assessed any of the 23 EM-specific milestone sub-competencies.²² Prior to reviewing narrative assessments, the study authors SC, AJ, and JR met to establish a shared mental model by reading and discussing the ACGME milestones framework. We reviewed the first 50 WBNA, resolved discrepancies as a group, and developed a guide detailing our interpretations of the milestones (Appendix A). We reviewed subsequent blocks of 50 WBNA, discussed discrepancies, and updated our guide in an iterative manner. This process

continued until the inter-rater reliability between the three reviewers was good ($k > 0.8$). We then independently reviewed the remaining WBNA and recorded the sub-competencies that each WBNA assessed in spreadsheets that included all 23 EM-specific milestone sub-competencies. For any given WBNA, there was no limit to the number of sub-competencies that could be assessed (Table 1).

Analysis

We performed descriptive statistics and conducted chi-square and analysis of variance tests for comparison on milestone proportions and means across three faculty levels to determine whether milestone sub-competencies were reported more frequently by specific faculty cohorts based on years of experience. All two-tailed significance tests were computed in Stata 13 with a set to 0.05 (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP).

RESULTS

During the 2016-2017 study period, we analyzed 2517 WBNA. WBNA were completed for 51 PGY 2-4 residents by 61 faculty members. Each resident received an average of 49 WBNA (range: 37 to 71), and each faculty member completed an average of 41 WBNA (range: 1 to 102). From the 2517 WBNA, we identified a total of 5130 milestone sub-competencies, with an average of 2.04 milestone sub-competencies assessed per WBNA. Of the 23 EM milestone sub-competencies, those most frequently identified through WBNA were multitasking, medical knowledge, practice-based performance improvement, patient-centered communication, and team management. The sub-competencies least frequently identified through WBNA were pharmacotherapy, airway management, anesthesia and acute pain management, goal-directed focused ultrasound, wound management, and vascular access. Overall, the frequency with which WBNA assessed individual sub-competencies was low, with 14 of the 23 sub-competencies being assessed in less than 5% of WBNA (Range: 0-33.3%) (Figure).

Junior faculty, defined as attending physicians in practice for less than five years, represented only 16% of the faculty, but completed 24% of the WBNA. Conversely, senior faculty,

defined as attending physicians in practice for greater than 15 years, represented 26% of the faculty, but completed only 17% of the WBNA. Mid-career faculty, defined as attending physicians in practice between 5-15 years, accounted for the 57% of the faculty and 58% of the WBNA. On average, junior faculty members identified 2.30 milestone sub-competencies per WBNA, compared to 2.03 milestone sub-competencies per WBNA for mid-career faculty, and 1.88 milestone sub-competencies per WBNA for senior faculty.

There was a statistically significant difference in milestone sub-competencies identified by faculty cohorts based on years of experience (ie, junior, mid-career, senior) in 14 of the 23 EM milestone sub-competencies (Table 2). On average, senior faculty tended to identify fewer individual milestone sub-competencies on WBNA when compared to their junior or mid-career faculty colleagues.

DISCUSSION

In this evaluation of 2517 WBNA at a single residency program, we found that each WBNA on average identified two milestone sub-competencies, with WBNA clustering around five specific sub-competencies and largely ignoring 60% of the sub-competencies. All sub-competencies are directly observable in the clinical environment,²² and while certain sub-competencies such as medical knowledge can be observed elsewhere (eg, in-service examination, mock oral boards), those related to patient care and procedural skills that are best observed in the clinical environment were notably absent from the WBNA. While junior faculty narratives assessed slightly more sub-competencies than mid-career or senior faculty narratives, the overall frequency of addressing milestone sub-competencies through WBNA was low.

These findings are concerning because when WBNA fail to comment on the majority of milestone sub-competencies, program directors and CCCs are left to make judgments regarding resident performance on a wide range of unassessed skills.⁸ This is detrimental to resident education, as the assumption of competence limits future targeted observations and interventions by faculty members, and it may either suppress a resident’s desire to self-report areas of weakness or it may promote a sense of inappropriate overconfidence when true performance lags behind resident self-assessment.

Despite the value of WBNA as an assessment tool,⁷⁻⁹ we hypothesize a lack of consistent faculty development as one reason for faculty’s poor performance in identifying numerous milestone sub-competencies. According to van der Vleuten et al.,²³ as an instrument seeks to assess higher levels on Miller’s pyramid,²⁴ the validity is more dependent on the assessors and the quality of the implementation as opposed to the instrument itself. Workplace-based observation and assessment of resident performance, and the subsequent narrative documentation of these observations and interpretations, is a skill that requires both training and practice. Too often, assessors receive little to no training in the practice of delivering WBNA,⁶ even though

Table 1. Example of a workplace-based narrative assessment identifying emergency medicine milestone sub-competencies.

Workplace-Based Narrative Assessment	Milestone(s) Assessed
“She has a great fund of knowledge, advocates for her patients, and does a great job managing the critically ill.”	Medical Knowledge Patient Centered Communication Emergency Stabilization
I love working with him. He’s very friendly, humble, hard-working, and enjoys learning.”	None applicable

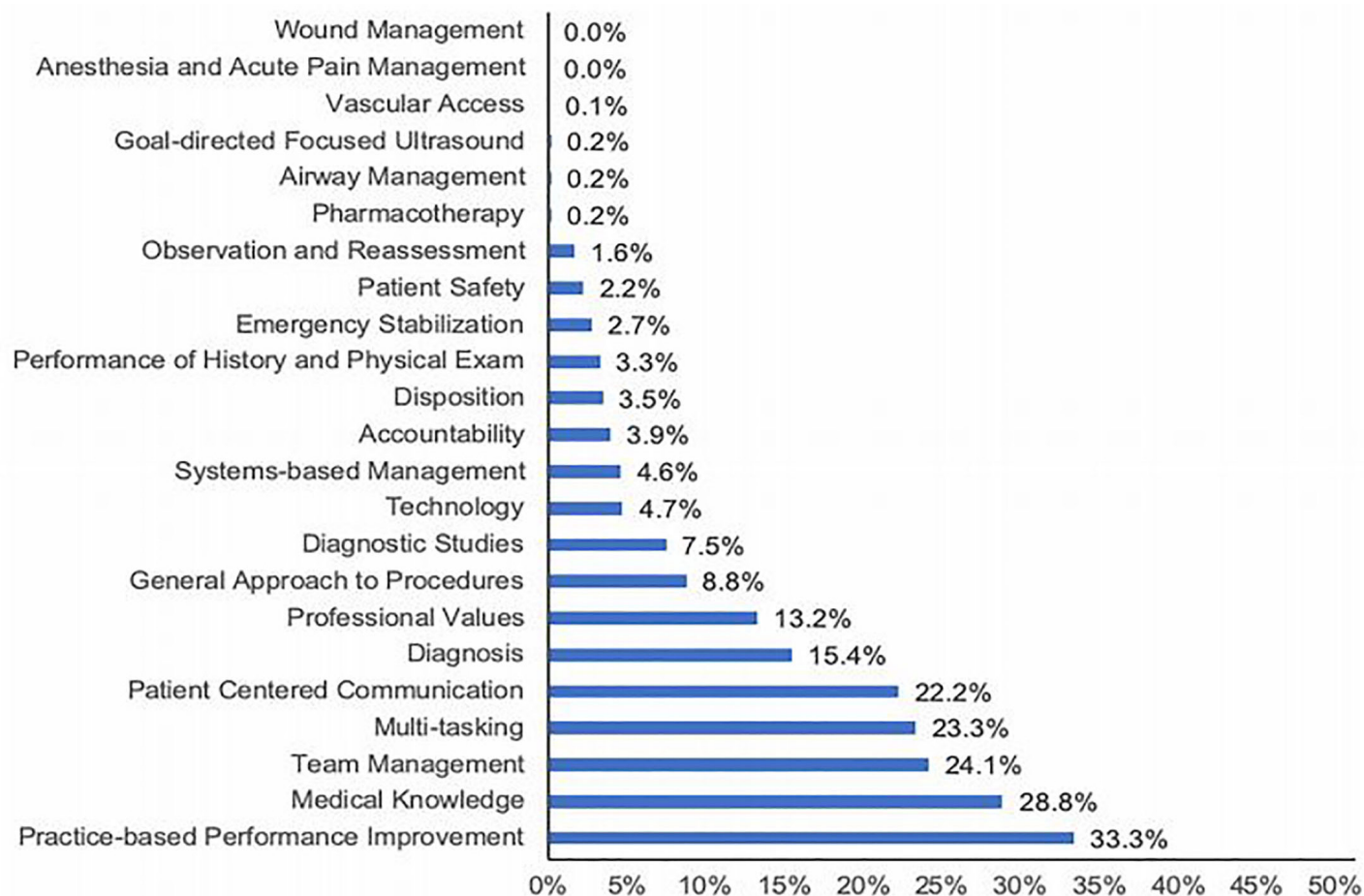


Figure. Frequency of emergency medicine milestone sub-competencies identified per workplace-based narrative assessment.

studies suggest that faculty development can improve the number and quality of narrative comments.⁹ While no recurrent faculty training program exists within our department, we do not know whether prior faculty development initiatives had been implemented in past years. Regardless, the fact that junior faculty narratives assessed slightly more sub-competencies than mid-career or senior faculty narratives suggests that if there were prior faculty development initiatives, they did not have a lasting effect.

Faculty were more likely to comment on sub-competencies relating to the ACGME core competencies of interpersonal and communication skills, practice-based learning and improvement, and medical knowledge, than on sub-competencies relating to patient care and procedural skills. This differs from prior studies conducted in general surgery and internal medicine training programs, which demonstrated a higher frequency of faculty comments regarding patient care and greater variability in comments regarding interpersonal communication skills.^{25,26} The difference in ACGME core competencies identified through WBNA between our study and prior studies may be due to differing specialty-specific, faculty-resident dynamics,

varying prompts and constructs of the WBNA used, faculty training in workplace-based assessment, or cultures inherent to respective specialties or institutions.

We found that junior faculty completed WBNA more frequently than senior faculty and their WBNA identified milestone sub-competencies more frequently. It is unclear why this may be; however, one explanation may be that junior faculty members are more familiar with the milestones framework than senior faculty. Additionally, junior faculty generally work more shifts than senior faculty, and thus may be more likely to observe and comment on observed behaviors.

LIMITATIONS

Our study had several notable limitations. It is a single-center, specialty-specific study, which limits its generalizability. The fact that we did not account for the number of shifts worked per faculty member limited our ability to assess for whether shift count influenced the differences among faculty cohorts. We did not account for faculty members who provided the same verbatim written commentary for each WBNA, regardless of resident

Table 2. Frequency of emergency medicine (EM) milestone sub-competencies identified by respective faculty assessor cohorts.

Emergency Medicine Milestone Sub-Competencies	Junior (<5 years)	Mid-Career (5-15 years)	Senior (>15 years)	p-value
Observation and reassessment	0.91%	2.54%	0.00%	<0.001
Disposition	2.28%	5.30%	0.16%	<0.001
Patient safety	4.33%	2.27%	0.64%	<0.001
Systems-based management	8.88%	4.81%	0.96%	<0.001
Diagnostic studies	12.07%	8.73%	1.44%	<0.001
Diagnosis	17.08%	17.47%	9.29%	<0.001
General approach to procedures	2.73%	5.23%	21.31%	<0.001
Patient-centered communication	14.12%	24.35%	23.08%	<0.001
Practice-based performance improvement	46.01%	27.03%	38.78%	<0.001
Multitasking	20.05%	25.86%	19.71%	0.002
Team management	30.07%	23.66%	20.99%	0.002
Medical knowledge	35.76%	27.10%	27.72%	0.002
Professional values	18.22%	12.24%	12.02%	0.003
Accountability	5.92%	3.78%	2.72%	0.028
Performance of history and physical exam	2.05%	4.06%	2.56%	0.056
Vascular access	2.00%	1.00%	0.00%	0.073
Airway management	0.68%	0.14%	0.16%	0.109
Emergency stabilization	3.87%	2.68%	1.92%	0.155
Goal-directed focused ultrasound	0.46%	0.14%	0.00%	0.176
Pharmacotherapy	0.00%	0.28%	0.00%	0.231
Technology	3.87%	5.23%	4.01%	0.325
Anesthesia and acute pain management	0.00%	0.00%	0.00%	-
Wound management	0.00%	0.00%	0.00%	-

performance, which was an observed practice. While this type of behavior may alter the overall frequencies of milestone sub-competencies our study identified, as well as the quality of the assessment provided, we chose to include their data because both residents and the CCC receive their comments on an individual level.

The WBNAs used at our institution did not include any prompting for faculty to comment on specific milestone sub-competencies, which may have resulted in lower frequencies of milestone sub-competencies identified. As a result, these findings may not be generalizable to institutions that use evaluation structures with specific milestone prompts. In addition, the lack of an annual formalized faculty training raises the question as to whether consistent faculty development would improve the frequency of milestone sub-competencies identified.

Finally, while we interpreted the WBNAs based on definitions and consensus, evaluators often “read between the lines” of narratives when providing summative assessments on residents.^{8,27} Therefore, identifying the frequency of milestone sub-competencies may undervalue the role of WBNAs in providing information for summative assessments.

However, we would contend that a more analytical process than evaluator gestalt is necessary for improved reliability and validity in providing competency judgments on trainees.

We recognize that identifying milestone sub-competencies is not the only measure in determining the quality of an assessment. Similar to prior research,²⁶ many WBNAs commented on non-ACGME themes. While this study was not designed to evaluate these comments, prior research has demonstrated their value to both faculty²⁷ and residents.⁷ However, if assessment of individual sub-competencies is desirable,^{8,28} targeted faculty development activities can be implemented to enhance sub-competency identification. With improved assessments that target previously unaddressed milestone sub-competencies, CCCs and program directors will have better guidance towards providing summative assessments regarding resident performance.⁸

Future research should examine the effectiveness of these faculty development programs in improving the frequency of milestone sub-competencies identified, as well as evaluating for satisfaction of both residents and faculty members post-intervention. In addition, given the disparities in assessments of male and female residents,²⁹ it is important to examine the

role that gender (of both the assessor and trainee) plays in identifying which sub-competencies are identified through WBNAs. Finally, while we believe our coding guide is generalizable to other EM programs interested in mapping WBNAs to the milestones, it is possible that natural language processing, which aims to program machines to interpret human language,³⁰ could replace the need for manual assessment of WBNAs. Future research could look at the feasibility of natural language processing in the evaluation of WBNAs.

CONCLUSION

Our study demonstrates that unstructured WBNAs identify relatively few milestone sub-competencies. Faculty tend to assess similar sub-competencies related to interpersonal and communication skills, practice-based learning and improvement, and medical knowledge, while neglecting sub-competencies related to patient care and procedural skills. These findings can help shape faculty development programs designed to improve assessments of specific workplace behaviors and provide more robust data in the summative assessment of residents.

Address for Correspondence: David Diller, MD, LAC+USC Medical Center, Keck School of Medicine of University of Southern California, Department of Emergency Medicine, 1200 North State St, Room 1060, Los Angeles, CA 90033. Email: diller@usc.edu.

Conflicts of Interest: By the *WestJEM* article submission agreement, all authors are required to disclose all affiliations, funding sources and financial or management relationships that could be perceived as potential sources of bias. No author has professional or financial relationships with any companies that are relevant to this study. There are no conflicts of interest or sources of funding to declare.

Copyright: © 2020 Diller et al. This is an open access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) License. See: <http://creativecommons.org/licenses/by/4.0/>

REFERENCES

- Nasca TJ, Philibert I, Brigham T, et al. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051-6.
- Team NK. Exploring the ACGME core competencies. Available at: <https://knowledgeplus.nejm.org/blog/exploring-acgme-core-competencies/>. Accessed 1/25/2019.
- Carter WA. Milestone myths and misperceptions. *J Grad Med Educ*. 2014;6(1):18-20.
- Dehon E, Jones J, Puskarich M, et al. Use of emergency medicine milestones as items on end-of-shift evaluations results in overestimates of residents' proficiency level. *J Grad Med Educ*. 2015;7(2):192-6.
- Schott M, Kedia R, Promes SB, et al. Direct observation assessment of milestones: problems with reliability. *West J Emerg Med*. 2015;16(6):871-6.
- Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA*. 2009;302(12):1316-26.
- 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):1617-21.
- Hanson JL, Rosenberg AA, Lane JL. Narrative descriptions should replace grades and numerical ratings for clinical performance in medical education in the United States. *Front Psychol*. 2013;4:668.
- Hatala R, Sawatsky AP, Dudek N, et al. Using in-training evaluation report (ITER) Qualitative comments to assess medical students and residents: a systematic review. *Acad Med*. 2017;92(6):868-79.
- Ginsburg S, Eva K, Regehr G. Do in-training evaluation reports deserve their bad reputations? A study of the reliability and predictive ability of ITER scores and narrative comments. *Acad Med*. 2013;88(10):1539-44.
- Cohen G, Blumberg P, Ryan N, P. S. Do final grades reflect written qualitative evaluations of student performance? *Teaching and Learning in Medicine*. 1993;5(1):10-15.
- Frohna A, Stern D. The nature of qualitative comments in evaluating professionalism. *Med Educ*. 2005;39(8):763-8.
- Guerrasio J, Cumbler E, Trosterman A, et al. Determining need for remediation through postrotation evaluations. *J Grad Med Educ*. 2012;4(1):47-51.
- Sargeant J, Armson H, Chesluk B, et al. The processes and dimensions of informed self-assessment: a conceptual model. *Acad Med*. 2010;85(7):1212-20.
- Govaerts MJ, van der Vleuten CP, Schuwirth LW, et al. The use of observational diaries in in-training evaluation: student perceptions. *Adv Health Sci Educ Theory Pract*. 2005;10(3):171-88.
- Lefebvre C, Hiestand B, Glass C, et al. Examining the effects of narrative commentary on evaluators' summative assessments of resident performance. *Eval Health Prof*. 2018;163278718820415.
- Ginsburg S, van der Vleuten C, Eva KW, et al. Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. *Adv Health Sci Educ Theory Pract*. 2016;21(1):175-88.
- Ginsburg S, Gold W, Cavalcanti RB, et al. Competencies "plus": the nature of written comments on internal medicine residents' evaluation forms. *Acad Med*. 2011;86(10 Suppl):S30-34.
- Lye PS, Biernat KA, Bragg DS, et al. A pleasure to work with—an analysis of written comments on student evaluations. *Ambul Pediatr*. 2001;1(3):128-31.
- Kiefer CS, Colletti JE, Bellolio MF, et al. The "good" dean's letter. *Acad Med*. 2010;85(11):1705-8.

21. Canavan C, Holtman MC, Richmond M, et al. The quality of written comments on professional behaviors in a developmental multisource feedback program. *Acad Med*. 2010;85(10 Suppl):S106-9.
22. Beeson M, Christopher T, Heidt J, et al. The Emergency Medicine Milestone Project: A Joint Initiative of the Accreditation Council for Graduate Medical Education and the American Board of Emergency Medicine. Available at: <https://www.acgme.org/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf>. Accessed January 25, 2019.
23. van der Vleuten CP, Schuwirth LW, Scheele F, et al. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol*. 2010;24(6):703-19.
24. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63-7.
25. Park YS, Zar FA, Norcini JJ, et al. Competency evaluations in the Next Accreditation System: contributing to guidelines and implications. *Teach Learn Med*. 2016;28(2):135-45.
26. Tekian A, Borhani M, Tilton S, et al. What do quantitative ratings and qualitative comments tell us about general surgery residents' progress toward independent practice? Evidence from a 5-year longitudinal cohort. *Am J Surg*. 2019;217(2):288-95.
27. Ginsburg S, Regehr G, Lingard L, et al. Reading between the lines: faculty interpretations of narrative evaluation comments. *Med Educ*. 2015;49(3):296-306.
28. Ekpenyong A, Baker E, Harris I, et al. How do clinical competency committees use different sources of data to assess residents' performance on the internal medicine milestones? A mixed methods pilot study. *Med Teach*. 2017;39(10):1074-83.
29. Dayal A, O'Connor DM, Qadri U, et al. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med*. 2017;177(5):651-7.
30. Chary M, Parikh S, Manini A, et al. A review of natural language processing in medical education. *West J Emerg Med*. 2019;20(1):78-86.