

Faculty development for the evaluation system: a dual agenda

Kellee L Oller
Cuc T Mai
Robert J Ledford
Kevin E O'Brien

Department of Internal Medicine,
University of South Florida Morsani
College of Medicine, Tampa, FL, USA

Abstract: Faculty development for the evaluation process serves two distinct goals. The first goal is to improve the quality of the evaluations submitted by the faculty. Providing an accurate assessment of a learner's capabilities is a skill and, similar to other skills, can be developed with training. Frame-of-reference training serves to calibrate the faculty's standard of performance and build a uniform language of the evaluation. Second, areas for faculty professional growth can be identified from data generated from learners' evaluations of the faculty using narrative comments, item-level comparison reports, and comparative rank list information. This paper presents an innovative model, grounded in institutional experience and review of the literature, to provide feedback to faculty evaluators, thereby improving the reliability of the evaluation process, and motivating the professional growth of faculty as educators.

Keywords: faculty development, students, residents, performance data, evaluation

Introduction

The evaluation system is reliant on faculty assessments of learners along their developmental trajectory through medical school and residency.¹ Therefore, it is critical that the faculty provide consistent and effective assessments of the performance of medical students and residents, who we will refer to as learners. This guide can serve as a framework for medical educators to develop and implement a dual-purpose formative faculty feedback system. Given the constraints of geographic dispersion and clinical pressures of faculty, it would be ideal to schedule several sessions across time and space in order to increase participation. Requiring faculty to attend one faculty development session per year would ensure that they continue to receive timely teaching performance data and are reoriented to the goals of assessing learners.

Within the faculty development session for the evaluation system, two diverse agendas can be accomplished. Faculty will receive training to enhance the quality of the evaluations they submit of learners and will receive feedback on personal performance metrics to identify strengths and weaknesses to target areas for improvement. The first agenda is improving the quality of evaluations submitted by the faculty. Reliable observation and assessment of a learner is an acquired skill that requires the evaluator to possess cognitive understanding and use assessment language that may not be intuitive to faculty evaluators.² Simply familiarizing them with the evaluation instrument is not sufficient. Prior work has centered on frame-of-reference training to calibrate faculty assessments to the grading scale.³⁻⁷ Additionally, faculty can receive

Correspondence: Kellee L Oller
Department of Internal Medicine,
University of South Florida Morsani
College of Medicine, 5 Tampa General
Circle, HMT-750, Tampa, FL 33606, USA
Tel +1 813 844 4919
Fax +1 813 844 1934
Email koller@health.usf.edu

Table 1 Outline of pre-session preparation and agenda topics

| Faculty development framework |
|---|
| Pre-session preparation |
| Build individualized performance portfolios |
| Schedule formal sessions across time and geographic sites |
| Agenda 1: Improve the quality of faculty assessments of learners |
| Review the evaluation tool faculty complete of learners |
| Discuss quality assessment language for the narrative component |
| Discuss determination of RIME synthesis level |
| Promote self-reflection of evaluator biases |
| Agenda 2: Use performance data from learner ratings to promote faculty growth |
| Encourage faculty self-assessment |
| Provide compiled narrative comments from learner evaluations of faculty |
| Discuss high-value teaching behaviors |
| Review faculty rank and comparison reports to identify areas for faculty growth |

Abbreviation: RIME, reporter–interpreter–manager–educator.

instruction on quality assessment language or the vocabulary that should be used to build an effective narrative statement. This training develops a shared mental model that faculty use when assessing learners' knowledge, skills, and attitudes.

The second agenda, providing feedback to the faculty as educators, is accomplished using performance data generated by the learners' ratings. Providing this explicit face-to-face individualized feedback promotes meaningful dialog among faculty and supports the ultimate goal of nurturing self-awareness and positive change. Faculty educators find faculty development aimed at teaching effectiveness valuable and leads to changes in their teaching behaviors that are detectable at the learner level.⁸

This guide is organized into pre-session preparation, Agenda 1 topics, and Agenda 2 topics (Table 1).

Pre-session: preparing for the faculty development session

Build individualized portfolios

Individualized teaching portfolios should include performance data generated by the learners' ratings of the faculty member. Examples of documents within a portfolio include the anonymous report of all narrative comments by learners, item-level comparison reports of faculty to their peers, and grading tendencies of the evaluator (Table 2). These documents are customized for each faculty member; while they can see where they measure compared to others in the department, the names other than their own are de-identified on the comparison documents. These documents can be created using data from an electronic evaluation management system and compiled by an administrative assistant.

Table 2 Summary of individualized faculty performance documents provided in portfolio

| Document | Description |
|-------------------------------|---|
| Self-assessment form | Evaluation form faculty completes at session to rate their own teaching practices |
| Narrative evaluation comments | Compilation of anonymous free response comments submitted by learners regarding faculty member |
| Faculty rank list | Departmental list ranked highest to lowest based on summative rating scores on all evaluations, names other than individual are blinded |
| Comparison report | Item-level comparison of faculty members average score and range in comparison to the departmental average and range |
| Bias tendencies report | Compares the average score of each learner by the faculty evaluator with the average score assigned by all other evaluators |

Schedule a formal face-to-face session

If the documents are merely sent to a faculty member, will they understand the comparison reports? Will they reflect on their narrative comments in order to identify areas of improvement? Will they review the evaluation rubric in order to improve the quality of evaluations they are submitting? Face-to-face feedback sessions ensure that the full potential of the portfolio is achieved.

Formal feedback sessions have been shown to increase the use of reporter–interpreter–manager–educator (RIME) terminology, improve grade consistency, and lead to less grade inflation or leniency.⁷ Faculty development targeted at enhancing teaching effectiveness overall results in gains in teaching skills detectable at the learner level.⁸ If only written feedback on faculty performance is provided, effective change typically does not occur.^{9,10} In fact, there can be diverse interpretations of performance data. Faculty may focus on the positive responses and their successes without identifying and reflecting on their weaknesses.¹⁰ Alternatively, some faculty may experience emotions of defensiveness, anger, or denial in response to negative feedback and in isolation are unable to develop constructive reflection and strategies for improvement.¹¹ Facilitated reflection of performance in a peer group setting has been associated with deeper reflection and higher quality improvement plans.¹²

Agenda 1: improve the quality of faculty evaluations of the learners

Review the evaluation tool faculty complete of learners

The institution-specific evaluation instrument should be reviewed with the faculty. A discussion of the standard level

of performance required of the learner to achieve each rating is helpful to calibrate your faculty's observations of the performance of the learners. What qualifies as "meets expectations" versus "exceeds expectations"? Where should most on par, or satisfactory, learners fall in your evaluation rubric? To guide faculty, give specific examples of the learner's skills or behaviors at each level on the evaluation scale.

Discuss quality language for the narrative component

When properly completed, the faculty narrative assessment of the learner is the most valuable component and highly contributes to grade determination.¹³ Faculty should be provided with practical and explicit examples of comments that are both helpful and unhelpful in the evaluation process (Table 3). The intention is to create a narrative that addresses behaviors that are modifiable and critical to success, rather than personality that is central to the learner's self-image. Evaluators often avoid negative comments and tend to create filler comments by referring to the learner's likelihood of future success or enjoyment of the process. Evaluations that contain generalities such as "Susan will make a good doctor" or "Andrew was a pleasure to work with" do not aid in the creation of a narrative that will help the learner develop their clinical skills and should be avoided. Specifically, it is critical to provide faculty with assessment language that focuses on the learner's knowledge level, clinical skills (clinical reasoning, application of pathophysiology, presentation efficiency and effectiveness, documentation effectiveness), and attitude (work ethic, receptiveness to feedback).¹⁴ Our goal is to reassure faculty that candid feedback based on behaviors and skills is well received by learners overall.

Table 3 Examples of narrative assessment language for two learners at the level of a reporter

Weak narrative language

"Mr. Learner was a very strong student. He was punctual, enthusiastic, and inquisitive. It was a pleasure to work with him. I have no doubt he will do well on his future endeavors."

Strong narrative language

"Mr. Learner performed at a level appropriate for his training. He currently is at a reporter level but is transitioning to interpreter (inconsistently at this point in time). His medical knowledge demonstrated an understanding of basic physiology and pathology and this was reflected in his differential diagnoses. He is on target for his progression as a clinician. He was able to complete a history and physical with pertinent positive and negative questions. Examination skills were basic, but not deficient. He was eager and enthusiastic, a valuable member of the team. To improve, he should strive to fill gaps in knowledge and advance his clinical acumen with improvement in physical examination skills. He was a pleasure to work with."

Define the RIME assessment framework

RIME terminology has been used to varying degrees for learner assessment since its inception by Pangaro.³ When uniformly integrated into the narrative assessment, the faculty determination of the synthesis level of the learner is apparent across clinical activities. In the RIME framework, a learner at the reporter level can collect historical information, perform foundational physical examination maneuvers, and recognize normal from abnormal; however, they have difficulty with effectively organizing disparate information and determining relevance with fluidity. Learners then progress through interpreter and manager stages to the educator level. Learners at the educator level have a deeper and more complex understanding of the medical literature, can apply translational medicine at the bedside, and educate others. This step represents more than basic self-directed learning and is considered aspirational. This is the cumulative application of knowledge, skills, and attitude that allows for increasingly sophisticated medical reasoning and decision making.³

To calibrate faculty assessments during the session, exercises can be performed to analyze different learner presentations or documentation for synthesis level. In addition, promote discussion of where the majority of learners should be in this model for their level of training. For example, even the most superb third-year medical student learner is not expected to reach the educator level. Integrating the RIME vocabulary into all evaluations provides a uniform language and can allow for a longitudinal assessment of a learner's growth across time and clinical activities.⁶

Promote self-reflection of evaluator biases

The evaluation system is a human process, and several factors, or biases, can influence the assessment of even the same student for the same observed encounter among different evaluators.^{2,15} What can we do, if anything, to decrease these individual biases to promote an accurate and honest assessment of a learner's capabilities? Gingerich et al¹⁶ described three disparate, but not mutually exclusive, perspectives on bias in the evaluation system. Evaluators can be considered trainable, fallible, meaningfully idiosyncratic, or a combination of such. From the trainable perspective, consistency in assessments can be improved with frame-of-reference training. The system is also perceived to be fallible since it relies on human judgment, which is imperfect. Stereotypes, emotions, and motives are difficult to self-regulate. Finally, it may be beneficial to embrace the idiosyncrasies of the evaluation system. As learners are challenged by different clinical

environments and patient care scenarios, even conflicting observations can be legitimate and contribute to a diverse, yet accurate, global assessment of the learner.¹⁶

Electronic evaluation management systems have the capability to generate individualized data of faculty evaluation practices. Within these systems, reports can be generated based on how an individual faculty member grades each learner compared to the mean of other evaluators grading the same learner. Alternatively, reports can look at average grades of the evaluator alone to identify trends in leniency or stringency. The purpose of providing this information is to facilitate self-reflection of grading trends and biases at hand.

Agenda 2: use performance data from learner ratings to promote faculty growth

Encourage faculty self-assessment

Prior to distributing learner ratings and narrative comments to the faculty, consider utilizing a self-assessment tool to supplement the reflective process. After faculty evaluate their teaching practices and effectiveness, provide the learner ratings for comparison. Utilizing the same evaluation tool that learners submit of faculty for the self-assessment allows for a point-by-point comparison of faculty and learner perceptions. The practice of combining self-assessment with learner ratings can be more effective than either alone in guiding faculty to identify discrepancies in their perceived strengths and weaknesses.¹⁷

Provide narrative comments from learner evaluations of the faculty

Faculty members are provided a report of all narrative comments submitted by the learners during the prior academic year. Similar to learners, faculty should be encouraged to review their narrative comments for constructive criticism and observations rather than focus on praise or personality. Once strengths and weaknesses are identified, then strategies to modify teaching behaviors can be discussed.

Discuss high-value teaching behaviors

Torre et al identified teaching behaviors that were deemed most valuable by learners in an inpatient setting. These behaviors included providing mini-lectures, teaching clinical skills (such as interpreting electrocardiograms and chest radiography), giving feedback on case presentations and differential diagnoses, and bedside teaching.¹⁸ A useful exercise that faculty can do independently is to review all learner

narrative comments and identify what high- and low-value behaviors they have performed during encounters with learners. Explicit discussion of valuable teaching behaviors can reinforce desired skills, but more importantly can encourage self-reflection of those that are not desirable. In a systematic review of faculty development for teaching improvement, such programs were highly regarded by faculty and resulted in increased knowledge (teaching skills, educational principles), skills (changes in teaching behaviors were detectable on the learner level), and attitude of participants (positive outlook on faculty development and teaching).⁸ Institutions should strive to support a culture where faculty members are evaluated on these specific behaviors so that high-quality teaching behaviors are reinforced.

Provide faculty rank lists and comparison reports to motivate change

Learner evaluations of faculty have been demonstrated to be a reliable and valid measure of evaluating faculty performance.^{10,19} Rank lists and comparison reports can be generated from the electronic evaluation management system. Rank lists are created by averaging all learner rating domain scores, then sorting faculty from highest to lowest average score. These lists can be individualized for each faculty member so they can see where they rank to the department, with the remainder of the names blinded. Comparison reports are more detailed comparisons that show faculty their individual average score and range for each domain in comparison to those of their department (Figure 1). This allows for identification of item-level discrepancies and identifies focused areas for improvement. For the faculty member in Figure 1, an identified area for improvement (circled) would be providing learners with more direct feedback on taking a history and performing a physical exam. When the faculty member identifies this area to improve upon, a subsequent group discussion could focus on strategies to integrate that task into clinical teaching practices.

Researchers in social studies have shown improvements in work performance when employees were given their performance metrics in comparison to groups.²⁰ Studies that have explicitly utilized comparison data in facilitated feedback suggest that it is most beneficial for normalizing the performance data.^{10,17,21} The negative consequence of comparative evaluation is downward comparisons, where the faculty member focuses on their superiority and fails to identify their own areas needing improvement.¹⁰ To counteract such, the faculty needs to identify group performance goals. Using learner ratings of faculty with facilitated

3. States goals and objectives clearly

| | Average score | Minimum score | Maximum score | Number of evaluations |
|------------|---------------|---------------|---------------|-----------------------|
| Individual | 5 | 5 | 5 | 17 |
| Group | 4.65 | 1 | 5 | 604 |

4. Observed and critiqued my histories and physical exams

| | Average score | Minimum score | Maximum score | Number of evaluations |
|------------|---------------|---------------|---------------|-----------------------|
| Individual | 4.59 | 1 | 5 | 17 |
| Group | 4.44 | 1 | 5 | 604 |

5. Effectively coaches clinical reasoning

| | Average score | Minimum score | Maximum score | Number of evaluations |
|------------|---------------|---------------|---------------|-----------------------|
| Individual | 5 | 5 | 5 | 17 |
| Group | 4.72 | 1 | 5 | 604 |

Figure 1 An excerpt from comparison report of an individual faculty member.

Note: Circle represents a potential area for faculty growth.

feedback can be helpful to better understand opportunities for improvement in teaching.

Conclusion

Direct observation and assessment of medical student and resident learners is the best model to evaluate clinical competencies and ultimately grade our learners; therefore, it is worthwhile to invest time to advance the process. Faculty development has been identified as the rate-limiting step in the evolution of the competency-based medical education approach.¹ Faculty should be provided with feedback on the quality of the evaluations they submit as well as data on their performance as an educator. While performance data should be reviewed annually, alternating topics or performing exercises to improve the quality of evaluations they submit are fun ways to keep faculty interested in the process. This dual agenda faculty development model provides a practical framework that can translate into more effective evaluations from your faculty, increase self-awareness of evaluation tendencies, and facilitate self-directed changes in teaching behaviors.^{7,8,16}

Acknowledgments

The authors thank Matel Galloway for her administrative support, leading to the success of faculty feedback sessions at our institution. We would like to thank Hesborn Wao, Ph.D., Stephanie Peters, M.A., and Barbara Pearce, M.A., for their thoughtful review of the manuscript.

Author contributions

All authors contributed toward data analysis, drafting and revising the paper and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

- 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.
- Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med.* 2003;15(4):270–292.
- Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med.* 1999;74(11):1203–1207.
- Hemmer PA, Pangaro L. Using formal evaluation sessions for case-based faculty development during clinical clerkships. *Acad Med.* 2000;75(12):1216–1221.
- Durning SJ, Pangaro LN, Denton GD, et al. Intersite consistency as a measurement of programmatic evaluation in a medicine clerkship with multiple, geographically separated sites. *Acad Med.* 2003;78(Suppl 10):S36–S38.
- 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;16:6354.
- Hemmer PA, Dadekian GA, Terndrup C, et al. Regular formal evaluation sessions are effective as frame-of-reference training for faculty evaluators of clerkship medical students. *J Gen Intern Med.* 2015;30(9):1313–1318.
- Steinert Y, Mann K, Centeno A, et al. A systemic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach.* 2006;28:497–526.
- Beijaard D, De Vries Y. Building expertise: a process perspective on the development or change of teachers' beliefs. *Eur J Teach Educ.* 1997;20(3):243–255.
- Arreola RA. *Developing a Comprehensive Faculty Evaluation System.* 3rd ed. Bolton, MA: Anker Publishing; 2007.
- Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback and acceptance and use. *Adv Health Sci Educ.* 2006;13:275–288.
- Boerboom TB, Jaarsma D, Dolmans DH, Scherpbier AJ, Mastenbroek NJ, van Beukelen P. Peer group reflection helps clinical teachers to critically reflect on their teaching. *Med Teach.* 2011;33(11):e615–e623.
- Hemmer PA, Papp KK, Mechaber AJ, Durning SJ. Evaluation, grading, and use of the RIME vocabulary on internal medicine clerkships: results of a national survey and comparison to other clinical clerkships. *Teach Learn Med.* 2008;20(2):118–126.
- Pangaro L. Investing in descriptive evaluation: a vision for the future of assessment. *Med Teach.* 2000;22:478–481.
- Albanese MA. Challenges in using rater judgments in medical education. *J Eval Clin Pract.* 2000;6(3):305–319.
- Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: assessor cognition from three research perspectives. *Med Educ.* 2014;48:1055–1068.
- Stalmeijer RE, Dolmans DH, Wolfhagen IH, Peters WG, van Coppenolle L, Scherpbier AJ. Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Adv Health Sci Educ.* 2010;15(3):315–328.

18. Torre DM, Simpson D, Sebastian JL, Elnicki DM. Learning/teaching activities and high-quality teaching: perceptions of third year medical students during an inpatient rotation. *Acad Med.* 2005;80(10):950–954.
19. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar NJ. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Med.* 2004;9:971–977.
20. Siero FW, Bakker AB, Dekker GB, van den Burg MTC. Changing organizational energy consumption behavior through comparative feedback. *J Environ Psychol.* 1996;16:235–246.
21. van der Leeuw RM, Slootweg IA, Heineman MJ, Lombarts K. Explaining how faculty members act upon residents' feedback to improve their teaching performance. *Med Educ.* 2013;47:1089–1098.

Advances in Medical Education and Practice

Dovepress

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education

including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>