

# Better Together: Development and Implementation of Fellow Group Evaluations of Faculty

Zachary A. Reese<sup>1</sup>, Jessica T. Lee<sup>1,2</sup>, and Caitlin Clancy<sup>1,3</sup>

<sup>1</sup>Division of Pulmonology, Allergy, and Critical Care, Department of Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania; <sup>2</sup>Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, Pennsylvania; and <sup>3</sup>Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania

## ABSTRACT

**Background:** High-quality trainee evaluations of faculty are essential for meaningful faculty development and for improving the clinical learning environment. However, concerns about anonymity can limit usefulness of trainee evaluations, particularly in smaller programs, such as subspecialty fellowships.

**Objective:** To develop and implement a fellow-driven group evaluation process to enhance trainee confidentiality and generate high-quality feedback for pulmonary and critical care medicine faculty.

**Methods:** A novel process was developed for faculty evaluation and feedback consisting of quarterly, structured, fellow-led group evaluation sessions focused on collecting confidential, behaviorally oriented, actionable feedback for faculty. Upper-year fellow moderators utilized a standard format to structure discussion, generating strengths and areas for growth for each faculty member while explicitly asking for input from fellows with divergent perspectives. Moderators compiled anonymized session notes for the program director, who delivered feedback to individual faculty. After the first six sessions, an electronic survey was distributed to assess fellow perceptions of the group evaluation model.

(Received in original form February 8, 2023; accepted in final form April 5, 2023)

This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0. For commercial usage and reprints, please e-mail Diane Gern.

**Correspondence and requests for reprints should be addressed to** Zachary A. Reese, M.D., Division of Pulmonology, Allergy, and Critical Care, Department of Medicine, Hospital of the University of Pennsylvania, 3400 Spruce Street, 8th Floor Maloney Building, Philadelphia, PA 19104. E-mail: zachary.reese@penmedicine.upenn.edu.

This article has a related editorial.

This article has a data supplement, which is accessible from this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org).

ATS Scholar Vol 4, Iss 3, pp 354–361, 2023  
Copyright © 2023 by the American Thoracic Society  
DOI: 10.34197/ats-scholar.2023-0023IN

**Results:** Thirty-seven faculty members were evaluated in 11 group sessions over 42 months. Fellows rated group-generated feedback as more confidential, more specific, more accurate, more efficient, more actionable, and less biased when compared with individual written evaluations ( $P < 0.01$  for all categories).

**Conclusion:** The authors successfully developed and implemented a process for fellow-led group evaluation of faculty, designed to facilitate fellow confidentiality and enrich the quality of feedback. Fellows preferred the group evaluation process and perceived group-generated feedback more favorably compared with individual written evaluations.

**Keywords:**

humans; teaching hospitals; graduate medical education; fellowship; critical care

Trainee evaluations of teaching faculty are an important tool for maintaining a safe and effective learning environment (1–3). The Accreditation Council for Graduate Medical Education requires that programs ensure that trainees receive high-quality teaching and provide opportunities to evaluate faculty teaching, skills, supervision, and professionalism (4). Faculty rely on trainee evaluations to develop their teaching skills, and medical schools use faculty evaluations to inform decisions about faculty promotion and awards (1, 5). However, concerns about validity and usefulness limit the value of faculty evaluations for professional development and program improvement (6, 7). Trainees tend to rate teaching effectiveness very highly, leading to a ceiling effect that can inhibit faculty growth (8). With limited utility in numerical ratings, programs place increased emphasis on narrative evaluations, although these have been noted to lack specificity (9).

Barriers limiting the accuracy and usefulness of trainee evaluations include evaluation fatigue, lack of time, and reluctance to provide constructive feedback because of hierarchy constraints, fear of retribution, desire to preserve ongoing relationships, and institutional culture (10, 11). Anonymous evaluations

are associated with lower ratings, supporting the impact of confidentiality concerns and potential retribution on accuracy of evaluations (12). Ensuring confidentiality is especially challenging for smaller programs, including subspecialty fellowships, where faculty work with a small number of trainees.

In the Pulmonary and Critical Care Medicine fellowship at the Hospital of the University of Pennsylvania, our trainees noted concerns about the confidentiality of their evaluations of faculty on the Accreditation Council for Graduate Medical Education Resident/Fellow Survey and expressed reluctance to provide constructive criticism in written evaluations to program leadership. To address these concerns and improve the overall quality of fellow evaluations of faculty, we developed and implemented a group-based evaluation and feedback model, focused on providing anonymous, behaviorally oriented, actionable feedback to faculty. Some of the results of this study have been previously reported in the form of an abstract (13).

## METHODS

Beginning in March 2019, our fellowship program implemented a process for fellow-led group evaluation of faculty.

Hour-long group evaluation sessions were scheduled approximately quarterly during protected didactic time, and fellows from all training years were invited to attend. Sessions were held either in a private conference room with videoconferencing capabilities to include remote participants or completely remotely when required by coronavirus disease (COVID-19) gathering restrictions.

A junior faculty member facilitated the pilot session and subsequently trained two upper-year fellows as moderators for further sessions. This training included a review of techniques to elicit behaviorally based examples and strategies to mitigate bias, such as purposely eliciting divergent opinions and asking for clarification and context when initial comments contained gendered or otherwise stereotyped language (*see* Appendix E1 in the data supplement). Before a new moderator led a session on their own, an experienced moderator observed as the new moderator led a session and gave feedback to the new moderator after the session.

Moderators selected three to four faculty members to evaluate per session, identified through a review of recent rotation schedules and by means of fellow suggestions, which moderators solicited via e-mail 1 week before the evaluation session. Early career faculty and faculty with a need for additional or updated evaluations (e.g., for promotion or award consideration) were prioritized on the basis of input from the program director. Initial sessions focused on the evaluation of faculty from inpatient rotations, and later sessions expanded to include fellow clinic preceptors. Outpatient faculty tended to work with a higher number of fellows, but in a less concentrated manner.

At the beginning of each session, moderators announced the purpose of group evaluation sessions and reminded

participating fellows that all feedback was anonymous. Moderators also reminded fellows of additional options for feedback, including contacting the moderator after the session.

During the sessions, moderators used a standard structure to guide discussion and elicit strengths and areas for growth for each faculty member. Moderators prompted fellows to provide specific examples and focus on behaviors rather than qualities. They also used strategies to counteract groupthink, defined as premature consensus seeking and closure in highly cohesive groups (14), including explicit solicitation of diverging opinions. Examples of these discussions are included in the data supplement (*see* Appendix E2). When conflicting feedback arose, moderators asked follow-up questions to clarify faculty behaviors and context. For example, disparate feedback between upper-year fellows and first-year fellows could represent faculty growth over time. Moderators took contemporaneous notes during sessions, excluding identifying information about fellows. After each session, moderators compiled written summaries of session notes, identifying themes discussed for both strengths and areas for growth. When fellow feedback was conflicting, moderators included specific examples to help the faculty understand where their behaviors may differ in certain situations (Appendix E2). Moderators sent evaluation summaries to the program director, who distributed the feedback to individual faculty via e-mail. Additionally, the program director offered opportunities for follow-up through one-on-one meetings and/or clarification with the moderators.

The group feedback process did not take the place of already established individual numerical ratings and written

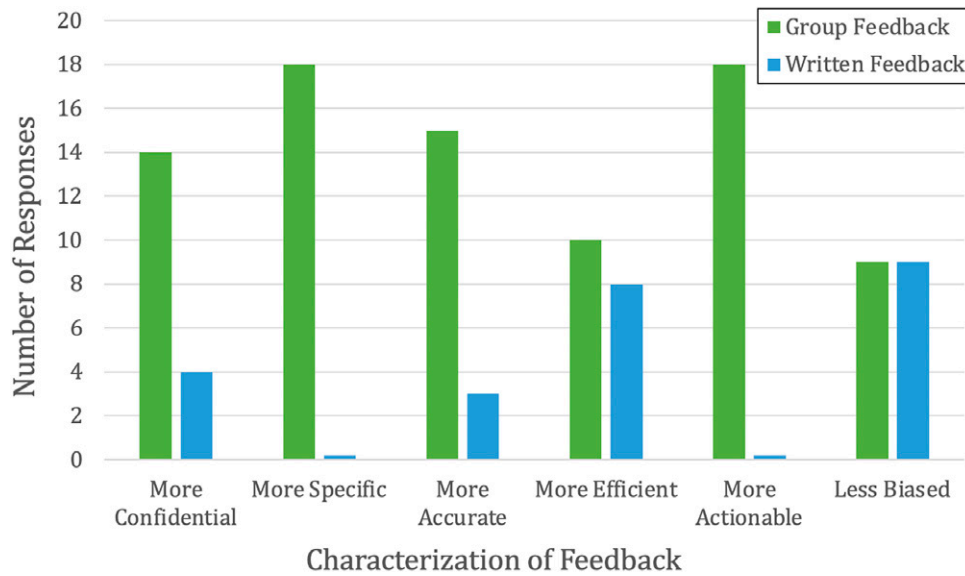


Figure 1. Direct comparison of group feedback versus written feedback.

narrative evaluations, as we felt it to be complementary to traditional methods of evaluation.

After the first six sessions, we distributed a seven-question electronic survey (Qualtrics) to all pulmonary and critical care medicine fellows to assess perceptions of accuracy, quality, and anonymity of feedback, with comparisons with the individual written evaluation process (see Appendix E3). We evaluated the survey responses

using paired *t* tests to compare mean Likert-scale ratings.

The institutional review board of the University of Pennsylvania determined this study to be exempt.

**RESULTS**

Over 42 months, we evaluated 37 unique faculty members in 11 group evaluation sessions. Nineteen of 28 fellows responded

Table 1. Comparisons of fellow ratings of group feedback versus written feedback

Feedback	Mean (95% CI)*		P Value
	Group Feedback Rating	Written Feedback Rating	
Confidential	4.4 (4.1–4.8)	3.0 (2.4–3.6)	0.002
Efficient	4.1 (3.6–4.5)	2.6 (1.9–3.3)	0.003
Higher quality	4.8 (4.6–5.0)	3.1 (2.6–3.6)	<0.001
Specific, behavioral based	4.5 (4.2–4.9)	2.1 (1.7–2.6)	<0.001
Should be continued	4.8 (4.6–5.0)	2.1 (1.6–2.5)	<0.001

Definition of abbreviation: CI = confidence interval.

\*Questions were asked on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Numeric means are reported.

to the survey (response rate, 68%), including 4 first-year fellows, 5 second-year fellows, 7 third-year fellows, and 3 fourth-year fellows. The majority of fellows ( $n = 18$ ; 95%) reported participating in at least one group evaluation session. Most fellows ( $n = 17$ ; 94%) rated group feedback as “much better” than written feedback.

When asked to compare methods, fellows rated feedback generated by the group evaluation as more confidential ( $n = 14$ ; 78%), more specific ( $n = 18$ ; 100%), more accurate ( $n = 15$ ; 83%), more efficient ( $n = 10$ ; 56%), more actionable ( $n = 18$ ; 100%), and less biased ( $n = 9$ ; 50%) (Figure 1). Mean Likert scale ratings were significantly higher for group feedback compared with written feedback along multiple domains, including confidentiality (4.4 vs. 3.0;  $P < 0.01$ ), efficiency (4.1 vs. 2.6;  $P < 0.01$ ), quality (4.8 vs. 3.1;  $P < 0.01$ ), and inclusion of specific, behaviorally based feedback (4.5 vs. 2.1;  $P < 0.01$ ). Fellow ratings of agreement that group evaluation sessions should continue were significantly higher compared with written evaluations (4.8 vs. 2.1;  $P < 0.01$ ) (Table 1). Examples of feedback generated from group evaluation sessions are included in the data supplement (Appendix E2).

## DISCUSSION

We successfully designed and implemented a fellow-led, group evaluation process that preserved fellow confidentiality and improved feedback for faculty. Fellows preferred these sessions to the written evaluation process and felt that feedback generated from group sessions was more confidential, more accurate, and more actionable. Additionally, fellows felt that the feedback they gave was stronger, more specific, and more behaviorally

oriented compared with written feedback. There was more equipoise over whether group feedback was less biased and/or more efficient when compared directly with individual feedback. This suggests that participants recognized the potential for bias in this method of gathering feedback.

Prior qualitative assessments of written evaluations of pulmonary and critical care faculty at our institution, including all learner levels, demonstrated that only a small subset (8%) contained a constructive comment, and written evaluations occasionally included inappropriate references to personal characteristics (e.g., appearance or humor) (15). With this fellow-led feedback innovation, every faculty member evaluated received a narrative evaluation with constructive feedback that did not include inappropriate feedback. Although we have not yet conducted an analysis of faculty feedback and reception to this process, multiple faculty members responded with appreciation for detailed feedback that was not present in written evaluations. Faculty have not relayed negative reactions through program leadership, and fellows have not reported any negative repercussions.

Although there are several published curricula on how faculty can provide effective feedback to medical trainees (1, 10, 16–21), there is a dearth of evidence on methods to improve trainee feedback of faculty. Pelsang and Smith have previously described group evaluation of faculty at the residency level (22). The authors of that study compared the numerical feedback score of faculty by residents individually and as a group, whereas our innovation evaluated trainee perceptions of the process and the quality of feedback. To our knowledge, this is the first assessment of group

evaluation of faculty in the fellowship setting.

### Limitations

There are several limitations to this innovation. First, we performed this innovation in a single, large fellowship program, and it is unclear whether our experience and findings are generalizable to fellowships at other institutions. Although fellows felt that the feedback generated was more anonymous than written feedback, this may be harder to accomplish in smaller fellowship programs. We scheduled the feedback sessions during a daily protected educational time, which helped to ensure high participation. Smaller fellowships, or those with less built-in educational time, may struggle to achieve a quorum sufficient for a rich, unbiased discussion. We only held the sessions quarterly, which limited the total number of faculty members discussed over the academic year. Although our intervention depended on the interest and leadership of upper-year fellows to moderate the sessions, we were able to train multiple moderators over the course of the program thus far. Even so, this fellow-driven intervention may be harder to sustain in a smaller or less-resourced fellowship.

We note several limitations of the group evaluation process itself. This study reports on fellow perceptions of feedback quality, and it does not capture the perceptions of other important stakeholders in the evaluation and feedback process (e.g., teaching faculty, program directors). Although we trained moderators to probe for dissenting opinions to minimize groupthink bias, the group format carries an inherent risk of

bias, as participants may have adjusted their feedback to adhere to group norms. There were a few occasions when an individual fellow reached out to the moderators separately to provide feedback that they did not share in the group setting. Additionally, because we invited fellows from all 4 years of fellowship to participate, feedback provided by upper-year fellows may have been outdated. Moderators were attuned to this possibility, and they framed the range of feedback as either evidence of faculty growth or as potential to work toward. They also sought out fellows with recent experiences working with the faculty member being evaluated to complement older feedback. Future plans include assessment of faculty perceptions of the group feedback model and qualitative assessment of the written feedback submitted during this time; there may be a spillover benefit to written individual feedback, as fellows participated in and observed their peers give specific, behaviorally oriented feedback. Identification of bias and further ways to mitigate it may improve the quality of the feedback.

### Conclusions

Implementation of a fellow-led group faculty evaluation process created opportunity for fellows to provide confidential feedback, which fellows perceived as more accurate and robust compared with individual written feedback. This innovation can be adapted by other fellowship programs as long as they have a peer moderator and a protected time in which to hold the sessions.

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

## REFERENCES

1. Williams BC, Pillsbury MS, Stern DT, Grum CM. Comparison of resident and medical student evaluation of faculty teaching. *Eval Health Prof* 2001;24:53–60.
2. Ramani S. Reflections on feedback: closing the loop. *Med Teach* 2016;38:206–207.
3. Telio S, Ajjawi R, Regehr G. The “educational alliance” as a framework for reconceptualizing feedback in medical education. *Acad Med* 2015;90:609–614.
4. Common program requirements (fellowship), version 3. Chicago, IL: Accreditation Council for Graduate Medical Education; 2022 [accessed 2023 Jan 30] Available from: [https://www.acgme.org/globalassets/pfassets/programrequirements/cprfellowship\\_2022v3.pdf](https://www.acgme.org/globalassets/pfassets/programrequirements/cprfellowship_2022v3.pdf).
5. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, *et al*. A review of the evaluation of clinical teaching: new perspectives and challenges. *Med Educ* 2000;34:862–870.
6. Fluit CRMG, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers: a systematic review of content and quality of questionnaires for assessing clinical teachers. *J Gen Intern Med* 2010;25:1337–1345.
7. Mariott D, Litzelman D. Teaching the teacher: is it effective? *Acad Med* 1998;73:S72–S74.
8. Albanese MA. Challenges in using rater judgements in medical education. *J Eval Clin Pract* 2000;6:305–319.
9. van der Leeuw RM, Schipper MP, Heineman MJ, Lombarts KMJMH. Residents’ narrative feedback on teaching performance of clinical teachers: analysis of the content and phrasing of suggestions for improvement. *Postgrad Med J* 2016;92:145–151.
10. Reddy ST, Zegarek MH, Fromme HB, Ryan MS, Schumann SA, Harris IB. Barriers and facilitators to effective feedback: a qualitative analysis of data from multispecialty resident focus groups. *J Grad Med Educ* 2015;7:214–219.
11. Ramani S, Könings KD, Mann KV, Pisarski EE, van der Vleuten CPM. About politeness, face, and feedback: exploring resident and faculty perceptions of how institutional feedback culture influences feedback practices. *Acad Med* 2018;93:1348–1358.
12. Afonso NM, Cardozo LJ, Mascarenhas OAJ, Aranha ANF, Shah C. Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. *Fam Med* 2005;37:43–47.
13. Reese Z, Lee JT, Clancy C, Kreider M. Evaluating the impact of pulmonary fellow group evaluation of faculty: a pilot study [abstract]. *Am J Respir Crit Care Med* 2021;203:A1540.
14. DiPierro K, Lee H, Pain KJ, Durning SJ, Choi JJ. Groupthink among health professional teams in patient care: a scoping review. *Med Teach* 2022;44:309–318.
15. Heath JK, Clancy CB, Carillo-Perez A, Dine CJ. Assessment of gender-based qualitative differences within trainee evaluations of faculty. *Ann Am Thorac Soc* 2020;17:621–626.
16. Ananthkrishnan S, Eyllon M, Noronha C. Can you hear me now? Helping faculty improve feedback exchange for internal medicine subspecialty fellows. *MedEdPORTAL* 2021;17:11099.
17. Fainstad T, McClintock AA, Van der Ridder MJ, Johnston SS, Patton KK. Feedback can be less stressful: medical trainee perceptions of using the Prepare to ADAPT (Ask-Discuss-Ask-Plan Together) framework. *Cureus* 2018;10:e3718.

18. Weinstein DF. Feedback in clinical education: untying the Gordian knot. *Acad Med* 2015;90:559–561.
19. Armson H, Lockyer JM, Zetkolic M, Könings KD, Sargeant J. Identifying coaching skills to improve feedback use in postgraduate medical education. *Med Educ* 2019;53:477–493.
20. Sargeant J, Lockyer JM, Mann K, Armson H, Warren A, Zetkolic M, *et al*. The R2C2 model in residency education: how does it foster coaching and promote feedback use? *Acad Med* 2018;93:1055–1063.
21. Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Med Teach* 2012;34:787–791.
22. Pelsang RE, Smith WL. Comparison of anonymous student ballots with student debriefing for faculty evaluations. *Med Educ* 2000;34:465–467.