



Enhancing Feedback on Professionalism and Communication Skills in Anesthesia Residency Programs

John D. Mitchell, MD,* Cindy Ku, MD,* Carol Ann B. Diachun, MD, MEd,† Amy DiLorenzo, MA,‡ Daniel E. Lee, MD, PhD,§ Suzanne Karan, MD,|| Vanessa Wong, BS,* Randall M. Schell, MD, MACM,‡ Marek Brzezinski, MD, PhD,¶ and Stephanie B. Jones, MD*

BACKGROUND: Despite its importance, training faculty to provide feedback to residents remains challenging. We hypothesized that, overall, at 4 institutions, a faculty development program on providing feedback on professionalism and communication skills would lead to (1) an improvement in the quantity, quality, and utility of feedback and (2) an increase in feedback containing negative/constructive feedback and pertaining to professionalism/communication. As secondary analyses, we explored these outcomes at the individual institutions.

METHODS: In this prospective cohort study (October 2013 to July 2014), we implemented a video-based educational program on feedback at 4 institutions. Feedback records from 3 months before to 3 months after the intervention were rated for quality (0–5), utility (0–5), and whether they had negative/constructive feedback and/or were related to professionalism/communication. Feedback records during the preintervention, intervention, and postintervention periods were compared using the Kruskal-Wallis and χ^2 tests. Data are reported as median (interquartile range) or proportion/percentage.

RESULTS: A total of 1926 feedback records were rated. The institutions overall did not have a significant difference in feedback quantity (preintervention: 855/3046 [28.1%]; postintervention: 896/3327 [26.9%]; odds ratio: 1.06; 95% confidence interval, 0.95–1.18; $P = .31$), feedback quality (preintervention: 2 [1–4]; intervention: 2 [1–4]; postintervention: 2 [1–4]; $P = .90$), feedback utility (preintervention: 1 [1–3]; intervention: 2 [1–3]; postintervention: 1 [1–2]; $P = .61$), or percentage of feedback records containing negative/constructive feedback (preintervention: 27%; intervention: 32%; postintervention: 25%; $P = .12$) or related to professionalism/communication (preintervention: 23%; intervention: 33%; postintervention: 24%; $P = .03$). Institution 1 had a significant difference in feedback quality (preintervention: 2 [1–3]; intervention: 3 [2–4]; postintervention: 3 [2–4]; $P = .001$) and utility (preintervention: 1 [1–3]; intervention: 2 [1–3]; postintervention: 2 [1–4]; $P = .008$). Institution 3 had a significant difference in the percentage of feedback records containing negative/constructive feedback (preintervention: 16%; intervention: 28%; postintervention: 17%; $P = .02$). Institution 2 had a significant difference in the percentage of feedback records related to professionalism/communication (preintervention: 26%; intervention: 57%; postintervention: 31%; $P < .001$).

CONCLUSIONS: We detected no overall changes but did detect different changes at each institution despite the identical intervention. The intervention may be more effective with new faculty and/or smaller discussion sessions. Future steps include refining the rating system, exploring ways to sustain changes, and investigating other factors contributing to feedback quality and utility. (*Anesth Analg* 2017;125:620–31)

From the *Department of Anesthesia, Critical Care and Pain Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts; †Department of Anesthesiology, University of Florida—Jacksonville, Jacksonville, Florida; ‡Department of Anesthesiology, University of Kentucky, Lexington, Kentucky; §Department of Anesthesiology and Pediatrics, University of California, San Diego, San Diego, California; ||Department of Anesthesiology, University of Rochester Medical Center, Rochester, New York; and ¶Department of Anesthesia and Perioperative Care, University of California, San Francisco, San Francisco, California.

Accepted for publication March 20, 2017.

Funding: Funded by a Research in Education Grant (2012–2015) from the Foundation for Anesthesia Education and Research (FAER, Schaumburg, IL).

The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (www.anesthesia-analgia.org).

Listen to this Article of the Month podcast and more from OpenAnesthesia.org by visiting <http://journals.lww.com/anesthesia-analgia/pages/default.aspx>.

Reprints will not be available from the authors.

Address correspondence to John D. Mitchell, MD, 1 Deaconess Rd, RB-470, Boston, MA 02215. Address e-mail to jdmitch@bidmc.harvard.edu.

Copyright © 2017 International Anesthesia Research Society
DOI: 10.1213/ANE.0000000000002143

Feedback is the process by which information on performance is conveyed back to those involved in the action or interaction. Feedback for medical education should be informed, objective, and not given for the purposes of evaluation.¹ Helpful feedback about suboptimal outcomes, which we will call “negative/constructive feedback,” is of especially high value in reflection on interactions and shaping future behaviors.

Feedback has always been an important element in helping residents achieve their full potential.¹ In the era of milestone-based performance assessment, feedback has taken on a central role in facilitating the progression of trainees to competency and beyond.² Training faculty in generating and delivering feedback is critical to successful provision of feedback to residents.³ Despite its importance, it has always been challenging for anesthesiology programs to provide such training. A 1999 survey by Rosenblatt and Schartel⁴ showed that only 20.2% of training programs at the time provided training in resident evaluation; a more recent

study showed that the number remains below 50%, with 90% of program directors desiring more training in feedback provision for their faculty.⁵ Across specialties, a variety of interventions have been tried but have met with varied success.^{1,6-8} Furthermore, the workflow, environmental, and time constraints endemic to the practice of procedural specialties make adoption of techniques from rounds or clinic-based specialties unlikely to be successful.^{2,9}

While all feedback is challenging to provide, professionalism and communication skills are among the topics on which providing feedback is most difficult for anesthesiologists.^{3,9-11} We therefore designed a faculty development program aimed at teaching faculty how to provide better feedback to residents, concentrating in the areas of professionalism and communication. Having piloted this program at Beth Israel Deaconess Medical Center, we refined the program and implemented it at 4 institutions nationally. These institutions, whose anesthesia residency programs ranged in size from 38 to 54 residents (postgraduate years 2-4) and whose faculty ranged in size from 40 to 75 attendings, shared a belief in the value of daily feedback in aiding resident development, but all struggled with faculty compliance and quality of feedback. Our hypothesis was that, overall, at the 4 institutions, this video-based faculty development program on provision of "difficult" feedback would lead to (1) an improvement in the quantity, quality, and utility of feedback provided to residents and (2) an increase in the percentage of feedback provided containing negative/constructive feedback as well as the percentage pertaining to professionalism and communication skills. To gain more understanding of our results, we explored these outcomes at the individual institutions as secondary analyses.

METHODS

Study Design and Participants

We implemented this study after obtaining institutional review board approval with a waiver of documentation of informed consent at all institutions. This before-after study was a prospective cohort study from the end of October 2013 to the beginning of July 2014 on feedback to residents from faculty in anesthesia at 4 institutions. Participants included anesthesia faculty and residents at all 4 institutions where faculty provide daily feedback to residents. One institution enrolled only faculty who were not exposed to our preliminary study; the other institutions enrolled all faculty. All anesthesia residents (postgraduate years 2-4) at the 4 institutions were included in the study as subjects of the feedback. Faculty and residents at each institution were recruited by the respective residency program director via e-mail.

We developed an educational program to raise awareness of the importance of feedback and to teach faculty how

to provide feedback. After implementing the program at the 4 institutions, we assessed feedback records at the 4 institutions for any changes.

Educational Program

Our educational program consisted of 2 discussion sessions for faculty and 1 discussion session for residents. Table 1 includes the total number of residents and faculty recruited as well as the attendance at these sessions overall and for each institution. Because of scheduling, not all faculty and residents attended the sessions in person; we therefore shared the materials with all the faculty and residents at the institutions through an e-mail sent by the residency program director so that they were able to review the materials at their own convenience.

Faculty Discussion Sessions

One investigator (J.D.M.) facilitated 2 discussion sessions for faculty at each institution. The discussion sessions focused on teaching faculty how to provide feedback, specifically feedback on professionalism and communication. We compiled a list of key points from these sessions and shared it with all faculty (see Supplemental Digital Content 1, Appendix 1, <http://links.lww.com/AA/B767>) along with the presentation slides from each session.

Faculty Session 1. The first session was a video-based discussion session that focused on basic feedback topics. The video that served as the basis for discussion simulated interactions between a faculty member and a resident who exhibits unprofessional behavior and poor communication skills (see <http://vimeo.com/22289080>, password: feedback). In the video, guidelines are provided on how faculty can address these issues and effectively provide feedback to the resident.

For the discussion session, we developed and used presentation slides (see Supplemental Digital Content 2, Appendix 2, <http://links.lww.com/AA/B768>) and discussion points from the video (see Supplemental Digital Content 3, Appendix 3, <http://links.lww.com/AA/B769>) to guide the session. We also used TurningPoint (Turning Technologies, LLC, Youngstown, OH), an audience response system, during the session to maintain audience engagement.

Faculty Session 2. The second faculty session was a follow-up discussion session that reviewed the feedback process and focused on examples of its application. It took place approximately 2 to 4 weeks after the first session. We developed and used presentation slides (see Supplemental Digital Content 4, Appendix 4, <http://links.lww.com/AA/B770>) to guide the session.

Table 1. Number and Percentage of Faculty and Residents Attending Sessions

Session	Institution 1 (N for Faculty = 7; N for Residents = 55)	Institution 2 (N for Faculty = 40; N for Residents = 51)	Institution 3 (N for Faculty = 76; N for Residents = 47)	Institution 4 (N for Faculty = 41; N for Residents = 37)	Total
Faculty session 1	4 (57.1%)	19 (47.5%)	19 (25%)	21 (51.2%)	63 (38.4%)
Faculty session 2	4 (57.1%)	7 (17.5%)	15 (19.7%)	23 (56.1%)	49 (29.9%)
Resident session	55 (100%)	18 (35.3%)	30 (63.8%)	18 (48.6%)	121 (63.7%)

Resident Discussion Session

At each institution, one investigator (S.B.J.) facilitated a resident discussion session at the same time as the first faculty session. The resident session, which aimed to ensure that residents were familiar with receiving feedback, focused on teaching residents how to solicit feedback and reinforced the value of feedback among the residents. We developed and used presentation slides (see Supplemental Digital Content 5, Appendix 5, <http://links.lww.com/AA/B771>) to guide the session and shared the slides with all residents.

Feedback Assessment

Data Collection. All 4 institutions had electronic systems in place to collect daily feedback from faculty to residents. Each day, faculty who worked with residents during the day received an e-mail asking them to complete an online form. The form asked the faculty member to provide feedback to the resident with whom he or she worked during the day. It included optional open-ended fields (sized for about 1–3 sentences each) in which the faculty member was asked to comment on the strengths and areas for improvement for the resident.

Using these existing systems, we collected open-ended feedback provided by faculty enrolled in the study at all 4 institutions from 3 months before faculty session 1 to 3 months after faculty session 2. We therefore had 3 data collection periods: preintervention (3 months before faculty session 1), intervention (time between the 2 sessions), and postintervention (3 months after faculty session 2). Data were collected concurrently at the 4 institutions (Figure 1). Because the sessions were scheduled based on available time and convenience for each institution, the intervention period was not uniform across all 4 institutions.

Rating System. Three experienced educators (J.D.M., C.K., S.B.J.) were identified as experts in feedback based on published criteria.¹² These experts developed a system to rate open-ended feedback based on published elements of high-quality feedback.^{1,4,5,13–15} These elements were converted to binary “yes/no” scoring. Three additional elements were included in the rating system: (1) whether or not the feedback contained negative/constructive feedback, (2) whether or not the feedback was related to professionalism/communication, and (3) the utility of the feedback. The final list of elements in the rating system was as follows:

1. Is the feedback detailed? (“Yes” = 1; “No” = 0)
2. Is the feedback specific? (“Yes” = 1; “No” = 0)
3. Is the feedback behavior-focused? (“Yes” = 1; “No” = 0)
4. Is the feedback destructive or harmful? (“Yes” = 0; “No” = 1)
5. Is the feedback actionable? (“Yes” = 1; “No” = 0)
6. Does the feedback contain negative/constructive feedback? (“Yes” = 1; “No” = 0)
7. Is the feedback related to professionalism/communication? (“Yes” = 1; “No” = 0)
8. As a program director or mentor, this feedback would help me devise a performance improvement plan for this resident. (“Strongly disagree” = 1, “Disagree” = 2, “Neither disagree nor agree” = 3, “Agree” = 4, “Strongly agree” = 5)

The experts met to discuss the scoring and develop specific definitions of each term; a criteria sheet was created for reference (see Supplemental Digital Content 6, Appendix 6, <http://links.lww.com/AA/B772>). To ensure consistent ratings, we decided that the raters would meet in person to discuss all the feedback records together and

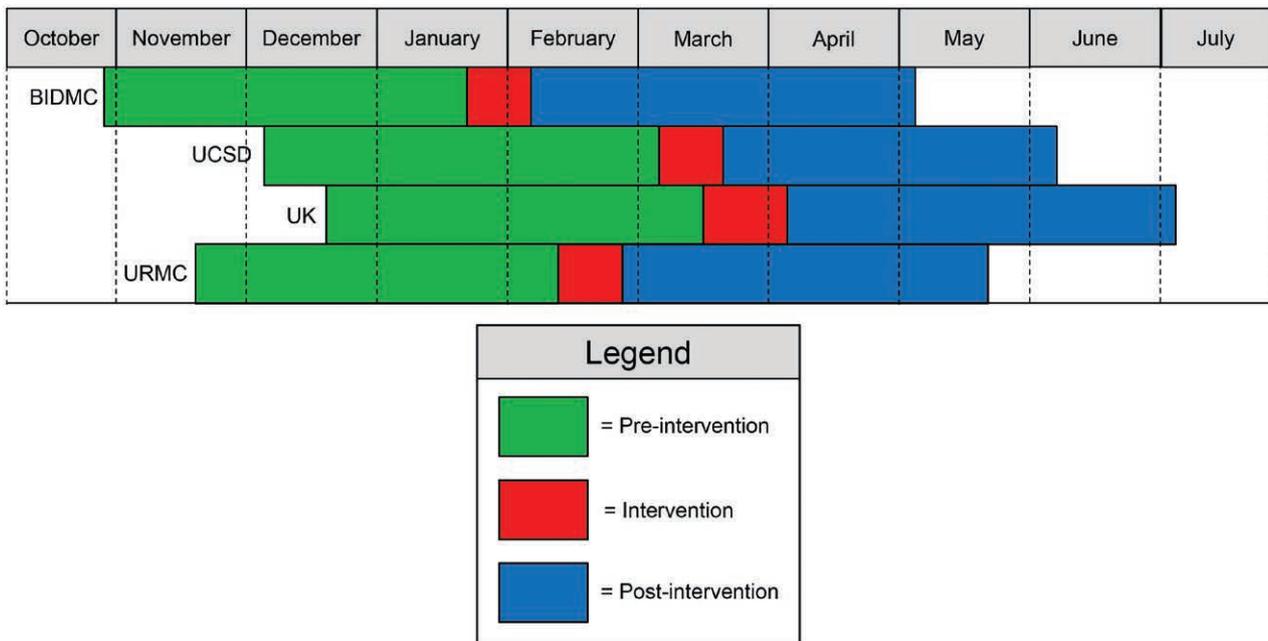


Figure 1. Data collection periods. Each institution had a 3-month preintervention period, a 2- to 4-week intervention period, and a 3-month postintervention period. BIDMC indicates Beth Israel Deaconess Medical Center; UCSD, University of California, San Diego; UK, University of Kentucky; URM, University of Rochester Medical Center.

unanimously agree on a score on each element for each feedback record.

After data collection was complete, the study coordinator (V.W.) combined all the feedback records from each institution and time period (preintervention period, intervention period, and postintervention period) and used Microsoft Excel (Microsoft Corporation, Redmond, WA) to assign a random number to each feedback record; records with repeated random numbers were assigned new random numbers until all records were assigned a unique random number. After sorting the records by their random numbers, the study coordinator gave only the open-ended responses to the 3 blinded experts in feedback (J.D.M., C.K., S.B.J.) to rate using the rating system described above. After scoring was complete, a principal component analysis (PCA) was performed to assess the factorial structure of the 5 elements measuring quality of feedback in the rating system (items 1–5). The PCA demonstrated that only 1 factor (eigenvalue: 2.07) had an eigenvalue of >1. These results suggested that only a single factor explained the majority of the variance in these elements. We also calculated a Cronbach α coefficient of .63 for these elements, a value that others in education research have classified as acceptable.¹⁶

Because the PCA demonstrated that only a single factor explained the majority of the variance in the 5 elements, we calculated a summative score by adding the scores for the 5 elements for each feedback record; this score represented the global rating for the quality of a feedback record. While questions 6 and 7 were not included in the global score for quality, they are characteristics of feedback related to our education program. Question 8 represented the utility rating for the feedback record.

Statistical Analysis

We performed the analyses described below using Microsoft Excel (Microsoft Corporation) and Stata (StataCorp LP, College Station, TX). With 5 primary hypotheses, we included a Bonferroni correction and considered a *P* value of less than .01 as significant when testing our primary hypotheses; for all other tests, we considered a *P* value of less than .05 as significant. Quality and utility ratings were not normal based on the Shapiro-Wilk test; attempts to transform the data were not successful, likely because of the ordinal nature of the variables, so we used nonparametric tests. Because each feedback record represented a different encounter/interaction depending on the faculty member giving the feedback, the resident receiving the feedback, and the case/clinical situation, we considered feedback records to be independent in our analyses. Data are reported as count, median (interquartile range), or percentage of feedback records.

The sample size (Table 2) was determined by the following:

1. Faculty schedules: For each day during the study, only faculty who were assigned to a resident were asked to provide feedback. Feedback records were available only from these faculty.
2. Provision of open-ended responses: Only feedback records with open-ended responses were included in our analyses.

Table 2. Sample Size
A. Number of Feedback Records

Time Period	Institution				Total
	1	2	3	4	
Preintervention (3 mo)	170	294	294	97	855
Intervention (2–4 wk)	36	42	89	8	175
Postintervention (3 mo)	143	336	371	46	896

B. Number of Faculty Providing Feedback

Time Period	Institution				Total
	1	2	3	4	
Preintervention (3 mo)	7	34	45	33	119
Intervention (2–4 wk)	7	11	23	7	48
Postintervention (3 mo)	7	33	49	18	107

C. Number of Residents Receiving Feedback

Time Period	Institution				Total
	1	2	3	4	
Preintervention (3 mo)	48	51	44	30	173
Intervention (2–4 wk)	21	25	37	8	91
Postintervention (3 mo)	46	49	47	25	167

3. Intervention schedule: Scheduling of the faculty and resident discussion sessions determined the start and end dates of the preintervention, intervention, and postintervention periods for each institution. As a result, it determined which (and how many) feedback records were included in each period. The intervention was scheduled based on available times at each institution.

With 855 records in the preintervention period and 896 records in the postintervention period, we had 85% power to detect a 0.2 mean change in the quality and utility ratings with a standard deviation of 1.4; we had 77% power to detect a change of 5% in the percentage of feedback records containing negative/constructive feedback and in the percentage of feedback records related to professionalism/communication.

Quantity. Overall, at the 4 institutions, we performed a χ^2 test to compare the proportion of feedback records requested that were provided in the preintervention period with the proportion in the postintervention period. We also calculated an odds ratio (OR; preintervention to postintervention) for the proportions. The intervention period was not included in this analysis. As part of our secondary analyses, we performed the same analysis for each institution.

Quality, Utility, and Other Characteristics. Overall, at the 4 institutions, we compared the following characteristics of feedback records during the preintervention, intervention, and postintervention periods:

1. Quality of feedback
2. Utility of feedback
3. Percentage of feedback containing negative/constructive feedback
4. Percentage of feedback related to professionalism/communication

The Kruskal-Wallis test was used for comparisons 1 and 2; the χ^2 test was used for comparisons 3 and 4. As part of our secondary analyses, we performed the same analyses for each institution.

Additional Analyses of Quality and Utility. To gain more insight on the quality and utility of feedback using our rating system, we performed the following exploratory analyses on the feedback records:

1. Using the Mann-Whitney *U* test, we compared the quality and utility of feedback records regardless of institution and time period:
 - a. Containing negative/constructive feedback versus not containing negative/constructive feedback
 - b. Related to professionalism/communication versus not related to professionalism/communication
2. We assessed the correlation between the quality and utility ratings, regardless of institution and time period, by calculating the Spearman rank correlation coefficient.
3. As a sensitivity analysis, we used generalized estimating equations to fit a model with a multinomial distribution and cumulative logit link function for the quality and utility ratings. Because the interaction between institution and time period was significant, we stratified the data by institution and ran the analysis for each institution taking into account the correlation between faculty and residents.

RESULTS

Quantity of Feedback

Table 3 depicts the feedback quantity at the institutions overall and individually for the preintervention and postintervention periods along with the OR. The quantity of feedback records during the postintervention period was:

1. Not significantly higher than during the preintervention period overall (preintervention: 855/3046 [28.1%]; postintervention: 896/3327 [26.9%]; OR: 1.06; 95% confidence interval [CI], 0.95–1.18; $P = .31$) and for institutions 2 (preintervention: 294/1726 [17.0%]; postintervention: 336/1811 [18.6%]; OR: 0.90; 95% CI, 0.76–1.07; $P = .24$) and 3 (preintervention: 294/598 [49.2%]; postintervention: 371/700 [53%]; OR: 0.86; 95% CI, 0.69–1.07; $P = .17$)

2. Significantly lower than during the preintervention period for institutions 1 (preintervention: 170/216 [78.7%]; postintervention: 143/264 [54.2%]; OR: 3.13; 95% CI, 2.08–4.69; $P < .001$) and 4 (preintervention: 97/506 [19.2%]; postintervention: 46/552 [8.3%]; OR: 2.61; 95% CI, 1.79–3.79; $P < .001$)

Quality of Feedback

Figure 2 summarizes the quality of feedback during the preintervention, intervention, and postintervention periods overall and at each institution. The institutions overall did not have a significant difference in quality ratings of feedback among the 3 time periods (preintervention: 2 [1–4]; intervention: 2 [1–4]; postintervention: 2 [1–4]; $P = .90$). Institution 1 had a significant difference in quality ratings among the 3 time periods with the median increasing from 2 (1–3) in the preintervention period to 3 (2–4) in the intervention period and remaining at 3 (2–4) in the postintervention period ($P = .001$). No other significant differences in the quality ratings were detected at the other institutions.

Figure 3 depicts the median (interquartile range) quality rating at 2-week intervals throughout the study for each institution. As shown in this figure, the median feedback quality rating was higher during the intervention and postintervention periods than during the preintervention period for institution 1.

Utility of Feedback

Figure 4 summarizes the utility of feedback during the preintervention, intervention, and postintervention periods overall and at each institution. The institutions overall did not have a significant difference in utility ratings of feedback among the 3 time periods (preintervention: 1 [1–3]; intervention: 2 [1–3]; postintervention: 1 [1–2]; $P = .61$). Institution 1 had a significant difference in utility ratings among the 3 time periods with the median increasing from 1 (1–3) in the preintervention period to 2 (1–3) in the intervention period and remaining at 2 (1–4) in the postintervention period ($P = .008$). No other significant differences in the utility ratings were detected at the other institutions.

Figure 5 depicts the median (interquartile range) utility rating at 2-week intervals throughout the study for each institution. As shown in this figure, the median feedback utility rating was higher during the intervention and postintervention periods than during the preintervention period for institution 1.

Other Characteristics of Feedback

Figure 6 summarizes the percentage of feedback records containing negative/constructive feedback and the percentage of feedback records related to professionalism/communication during the preintervention, intervention, and postintervention periods overall and at each institution.

Table 3. Feedback Quantity^a

Time Period	Institution 1	Institution 2	Institution 3	Institution 4	Overall
Preintervention	170/216 (78.7%)	294/1726 (17.0%)	294/598 (49.2%)	97/506 (19.2%)	855/3046 (28.1%)
Postintervention	143/264 (54.2%)	336/1811 (18.6%)	371/700 (53%)	46/552 (8.3%)	896/3327 (26.9%)
Odds ratio (preintervention to postintervention)	3.13 (2.08–4.69)	0.90 (0.76–1.07)	0.86 (0.69–1.07)	2.61 (1.79–3.79)	1.06 (0.95–1.18)

^aFeedback quantity is reported as the proportion (percentage) of the number of feedback records requested in the time period. Odds ratios are reported as odds ratio (95% confidence interval).

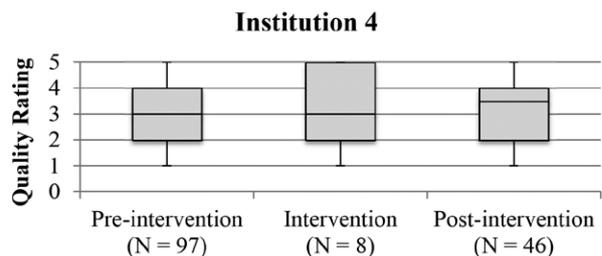
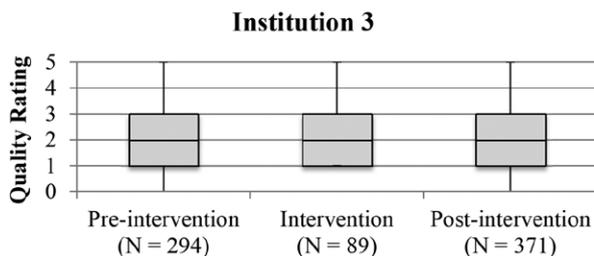
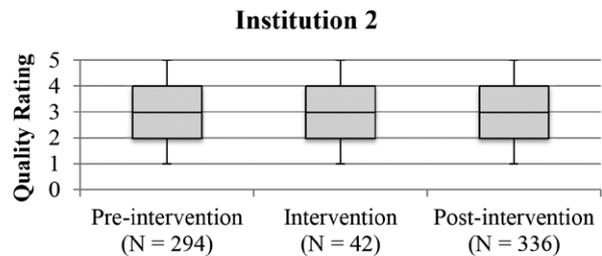
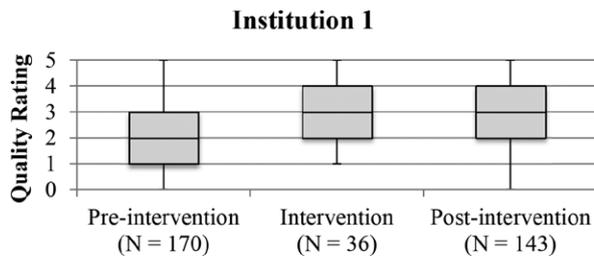
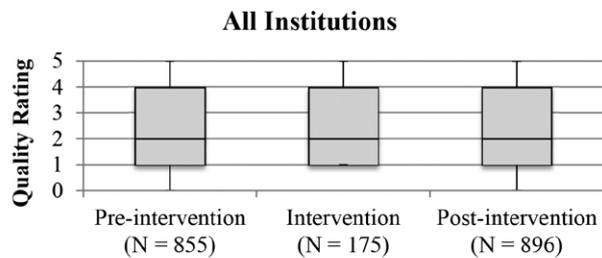


Figure 2. Quality of feedback during the preintervention, intervention, and postintervention periods. The institutions overall did not have a significant difference in quality ratings of feedback among the 3 time periods (preintervention: 2 [1–4]; intervention: 2 [1–4]; postintervention: 2 [1–4]; $P = .90$). Institution 1 had a significant difference in quality ratings among the 3 time periods with the median increasing from 2 (1–3) in the preintervention period to 3 (2–4) in the intervention period and remaining at 3 (2–4) in the postintervention period ($P = .001$). No other significant differences in the quality ratings were detected at the other institutions.

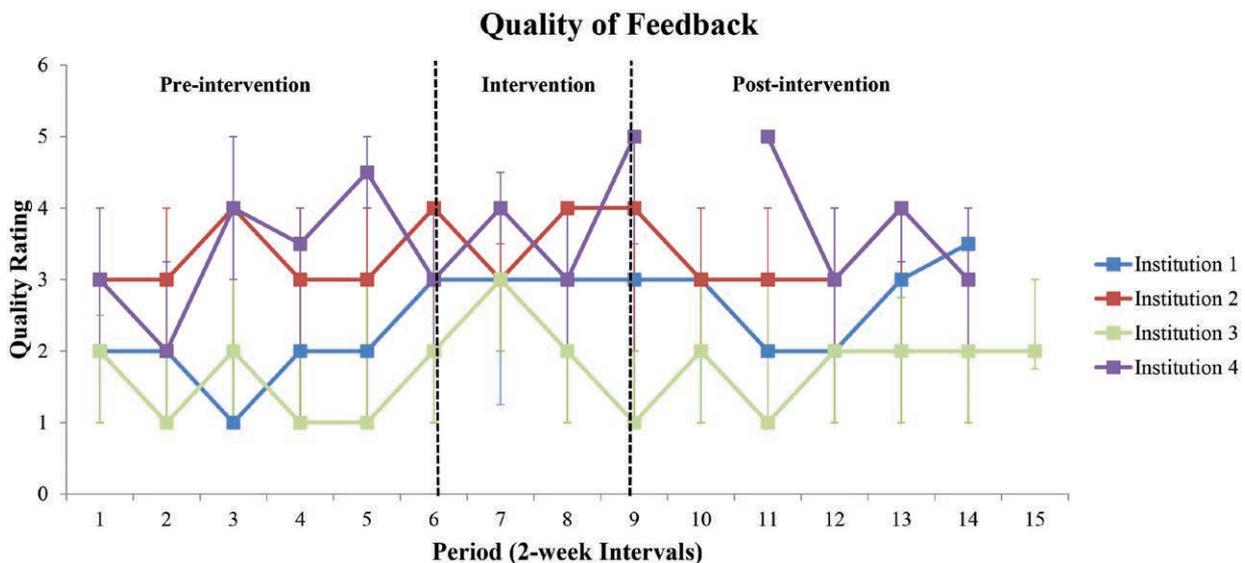


Figure 3. Median (interquartile range) quality rating at 2-week intervals throughout the study for each institution. The median feedback quality rating was higher during the intervention and postintervention periods than during the preintervention period for institution 1.

Negative/Constructive Feedback. The institutions overall did not have a significant difference in the percentage of feedback records containing negative/constructive feedback among the 3 time periods (preintervention:

27%; intervention: 32%; postintervention: 25%; $P = .12$). Institution 3 had a significant difference in the percentage of feedback records containing negative/constructive feedback among the 3 time periods with the percentage

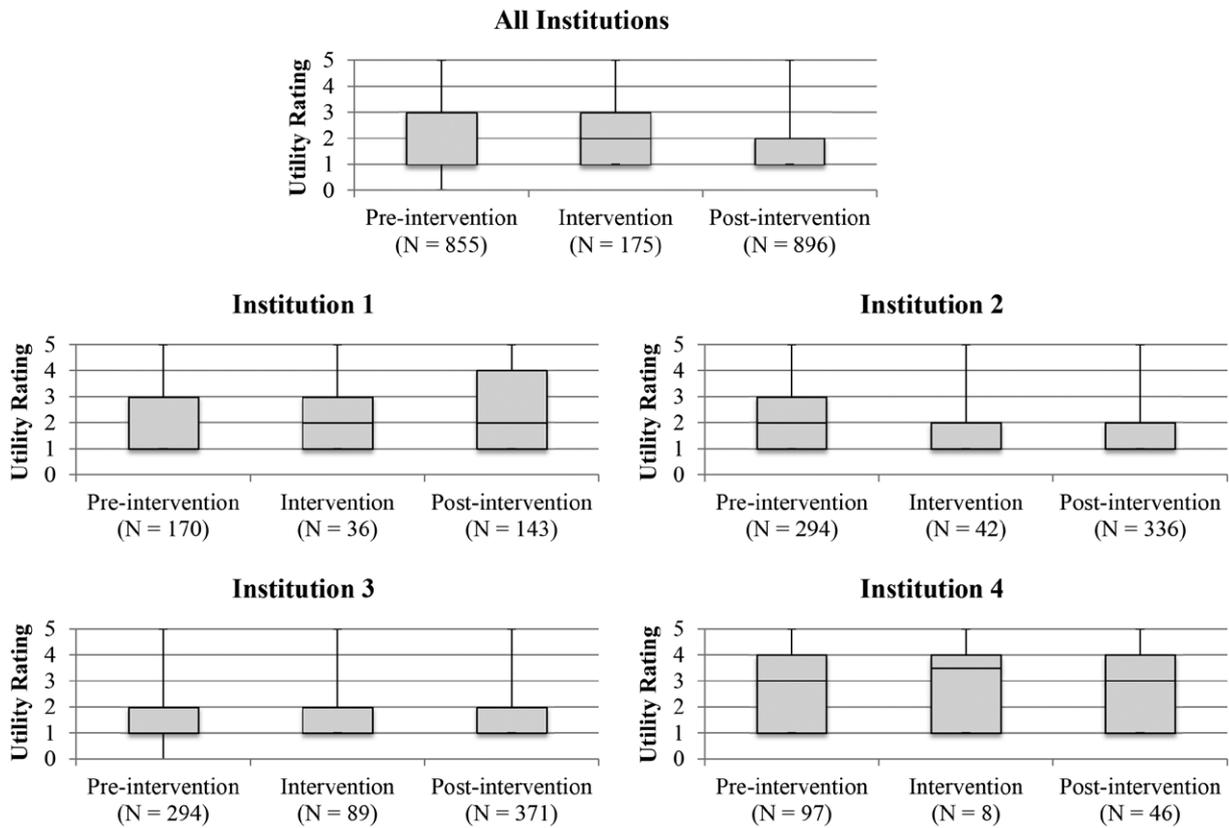


Figure 4. Utility of feedback during the preintervention, intervention, and postintervention periods. The institutions overall did not have a significant difference in utility ratings of feedback among the 3 time periods (preintervention: 1 [1–3]; intervention: 2 [1–3]; postintervention: 1 [1–2]; $P = .61$). Institution 1 had a significant difference in utility ratings among the 3 time periods with the median increasing from 1 (1–3) in the preintervention period to 2 (1–3) in the intervention period and remaining at 2 (1–4) in the postintervention period ($P = .008$). No other significant differences in the utility ratings were detected at the other institutions.

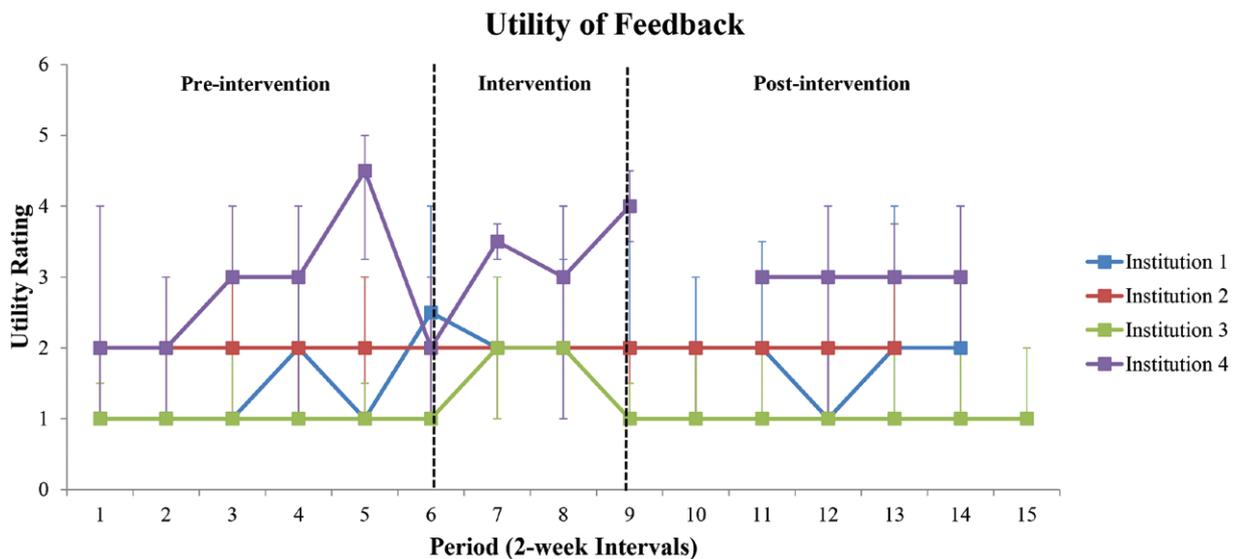


Figure 5. Median (interquartile range) utility rating at 2-week intervals throughout the study for each institution. The median feedback utility rating was higher during the intervention and postintervention periods than during the preintervention period for institution 1.

increasing from 16% in the preintervention period to 28% in the intervention period and decreasing to 17% in the postintervention period ($P = .02$). No other significant differences in this characteristic were detected at the other institutions.

Figure 7 depicts the percentage of feedback records containing negative/constructive feedback at 2-week intervals throughout the study for each institution. The percentage of feedback records containing negative/constructive feedback was highest at the beginning of the intervention period for institution 3.

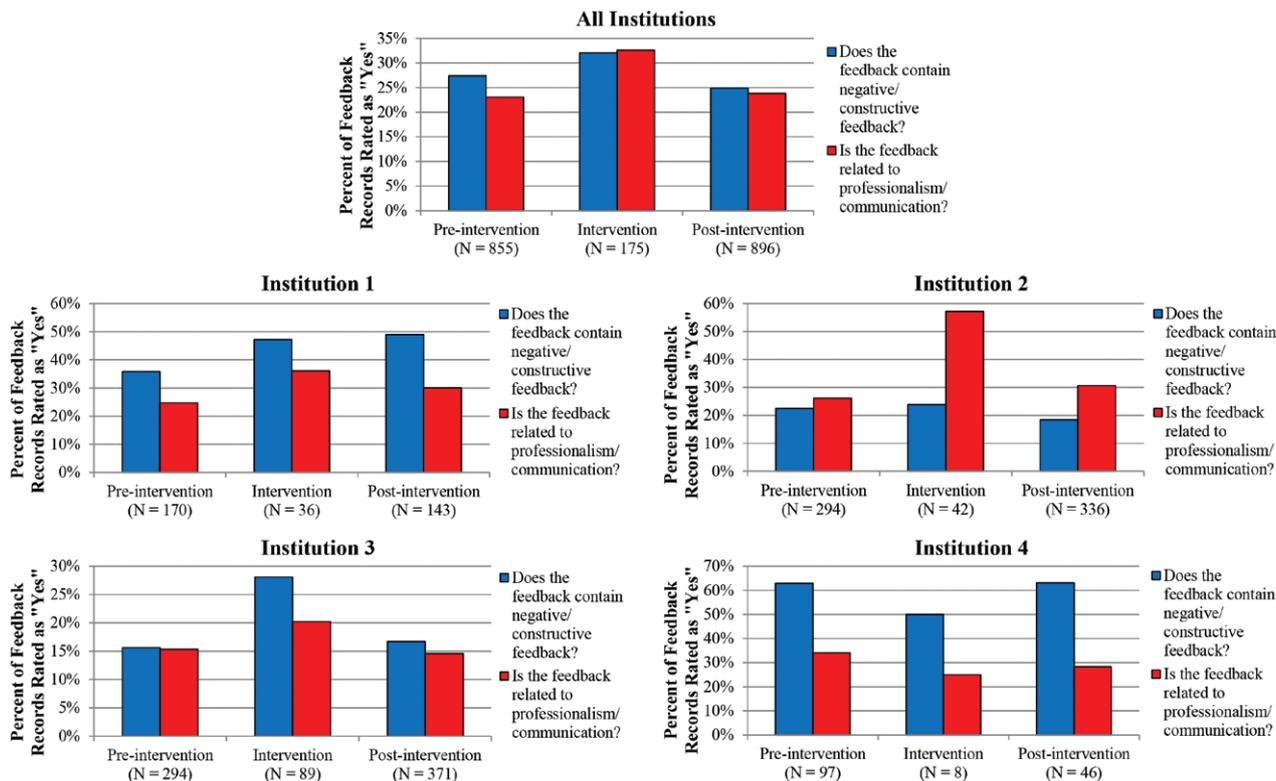


Figure 6. Negative/constructive feedback and feedback related to professionalism/communication during the preintervention, intervention, and postintervention periods. The institutions overall did not have a significant difference in the percentage of feedback records containing negative/constructive feedback ($P = .12$) or the percentage of feedback records related to professionalism/communication ($P = .03$) among the 3 time periods. Institution 3 had a significant difference in the percentage of feedback records containing negative/constructive feedback among the 3 time periods with the percentage increasing from 16% in the preintervention period to 28% in the intervention period and decreasing to 17% in the postintervention period ($P = .02$). Institution 2 had a significant difference in the percentage of feedback records related to professionalism/communication with the percentage increasing from 26% in the preintervention period to 57% in the intervention period and decreasing to 31% in the postintervention period ($P < .001$). No other significant differences in these characteristics were detected at the other institutions.

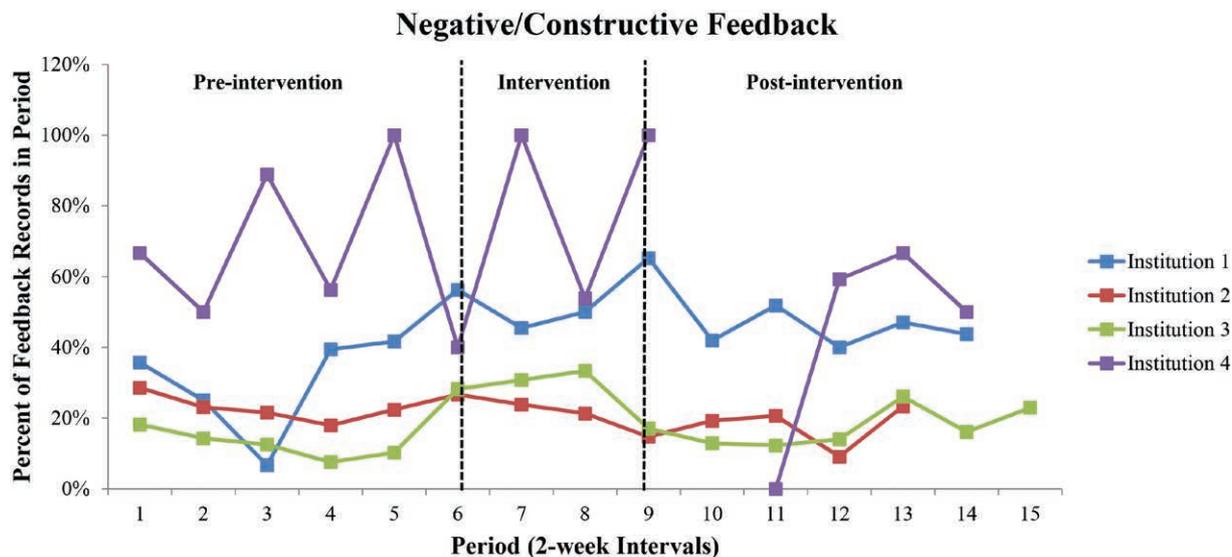


Figure 7. Percentage of feedback records containing negative/constructive feedback at 2-week intervals throughout the study for each institution. The percentage of feedback records containing negative/constructive feedback was highest at the beginning of the intervention period for institution 3.

Feedback Related to Professionalism/Communication. The institutions overall did not have a significant difference in the percentage of feedback records related to professionalism/communication among the 3 time periods

(preintervention: 23%; intervention: 33%; postintervention: 24%; $P = .03$). Institution 2 had a significant difference with the percentage increasing from 26% in the preintervention period to 57% in the intervention period and decreasing

to 31% in the postintervention period ($P < .001$). No other significant differences in this characteristic were detected at the other institutions.

Figure 8 depicts the percentage of feedback records related to professionalism/communication at 2-week intervals throughout the study for each institution. The percentage of feedback records related to professionalism/communication increased at the beginning of the intervention period for institution 2.

Additional Analyses of Quality and Utility

The quality and utility of feedback records containing negative/constructive feedback (quality: 4 [3–5]; utility: 4 [3–4]) were higher than those of feedback records not containing negative/constructive feedback (quality: 2 [1–3]; utility: 1 [1–2]; $P < .001$ for both quality and utility; Figure 9). Likewise, the quality and utility of feedback records related to professionalism/communication (quality: 4 [3–4]; utility: 3 [2–4]) were higher than those of feedback records not related to professionalism/communication (quality: 2 [1–3]; utility: 1 [1–2]; $P < .001$ for both quality and utility; Figure 9).

The Spearman rank correlation coefficient between the quality and utility ratings for the feedback records was 0.80 (95% CI, 0.79–0.82).

After taking into account the correlation between faculty and residents, there was a significant difference among the time periods in quality ($P < .001$) and utility ($P = .01$) ratings at institution 1. There were no significant differences overall or at the other institutions.

DISCUSSION

We developed and implemented an educational program to enhance the quantity, quality, and utility of daily feedback from faculty to residents in 4 unique academic anesthesiology departments. While no significant changes in feedback quantity, quality, and utility were detected overall after the intervention, one institution (institution 1) had significantly higher quality and utility ratings after the intervention.

Because this center, unlike the other centers, applied the intervention only to newly hired staff, the ideal time to implement the intervention may be when a faculty member is new to providing feedback. This finding is consistent with a study by Møystad et al,¹⁷ where novice teachers in dental education reported more benefits from a faculty development program than experienced teachers. Also, compared to the other centers, this center had a lower number of faculty at each discussion session (less than 10), which may have enhanced the interactive nature and effectiveness of the sessions and contributed to the improvement in feedback quality and utility.

Despite a focus in the educational program on providing negative/constructive feedback and feedback on professionalism and communication, the percentage of feedback records expressing these elements did not increase overall after the intervention. However, institutions 3 and 2 experienced significant increases in the percentage of records containing negative/constructive feedback and the percentage of records related to professionalism/communication, respectively, during the intervention, but these changes declined toward baseline in the 3 months following the end of the intervention. This result suggests that follow-up sessions may be necessary to sustain change. Review of the literature suggests that feedback involves a complex interplay of faculty, resident, and environment so that greater focus on departmental culture and resident receptivity to feedback may also sustain the transient improvements we noted.^{18–20}

In our study, feedback related to professionalism/communication and feedback containing negative/constructive feedback received higher quality and utility ratings. This result suggests that these 2 attributes of feedback may be markers of higher quality and more useful feedback. It may also suggest that our rating system for quality and utility is skewed toward weighing these 2 attributes more heavily than the other attributes. Future studies should investigate whether these 2 attributes are more prevalent in feedback from more experienced providers and whether emphasizing

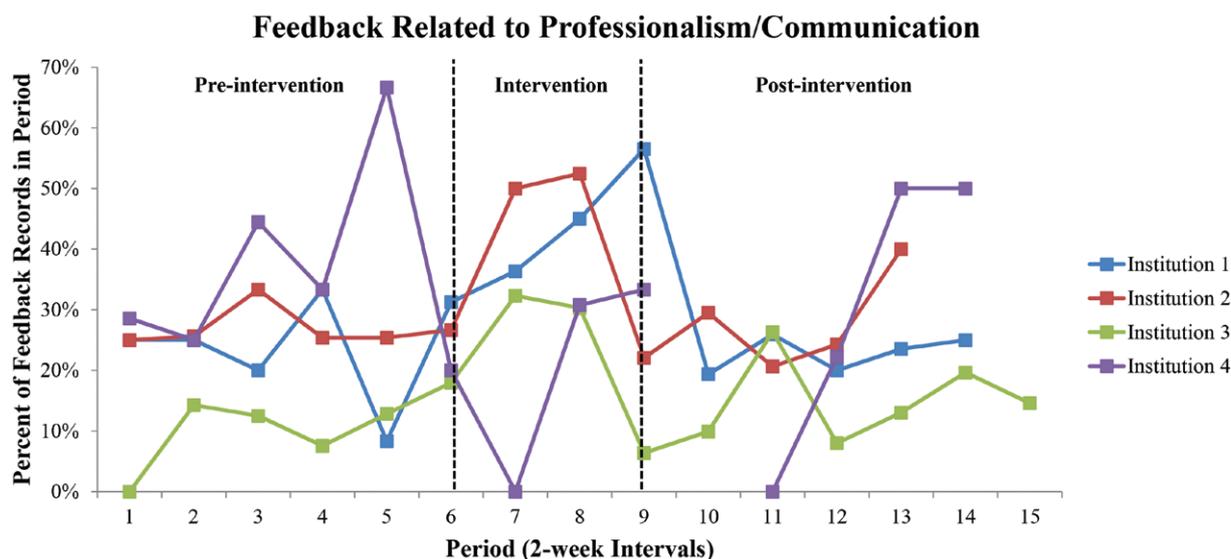


Figure 8. Percentage of feedback records related to professionalism/communication at 2-week intervals throughout the study for each institution. The percentage of feedback records related to professionalism/communication increased at the beginning of the intervention period for institution 2.

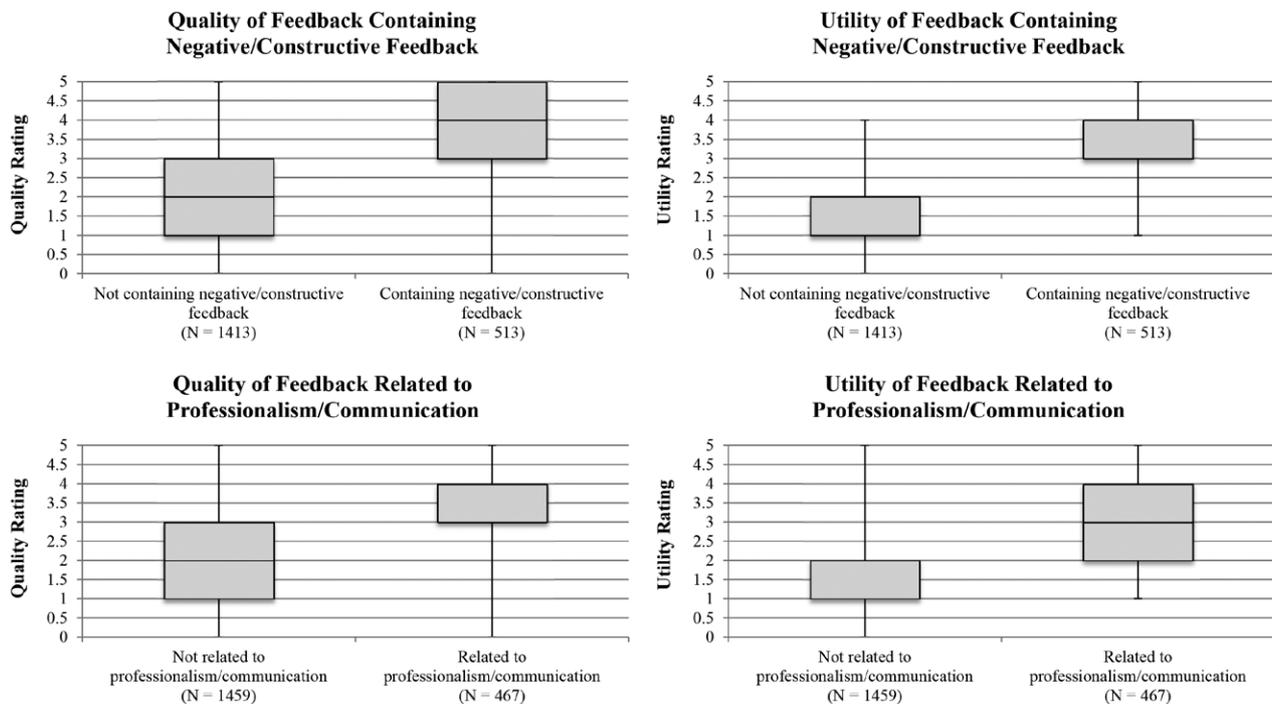


Figure 9. Quality and utility of feedback records containing negative/constructive feedback and feedback records related to professionalism/communication. The quality and utility of feedback records containing negative/constructive feedback (quality: 4 [3–5]; utility: 4 [3–4]) were higher than those of feedback records not containing negative/constructive feedback (quality: 2 [1–3]; utility: 1 [1–2]; $P < .001$ for both quality and utility). Likewise, the quality and utility of feedback records related to professionalism/communication (quality: 4 [3–4]; utility: 3 [2–4]) were higher than those of feedback records not related to professionalism/communication (quality: 2 [1–3]; utility: 1 [1–2]; $P < .001$ for both quality and utility).

these 2 attributes in faculty development programs would increase the quality and utility of feedback.

Finally, we found a strong positive correlation between the quality and utility ratings of feedback, suggesting that higher-quality feedback tends to be more useful to program directors and mentors. Thus, for feedback to be useful to program directors and mentors, faculty development programs should focus on teaching faculty how to provide higher-quality feedback.

Limitations and Strengths

We note the following limitations in this study, which have provided us guidance for future investigations:

1. Selection bias: Although all faculty were invited to participate in the study, faculty schedules determined whether they were included in the study. We did not collect demographic data on the faculty and therefore were not able to formally compare faculty who participated in the study to those who did not. However, because we included faculty from multiple institutions, we believe that selection bias did not have much influence on our results.
2. Participation in discussion sessions: All faculty were invited to participate in the discussion sessions, but, overall, only about one-third of the faculty were able to attend either session because of scheduling. We therefore shared the key points and slides from the sessions with all faculty. However, reviewing the materials alone is not the same as participating in the discussion sessions. It is possible that this

difference may have contributed to the different changes we detected at the institutions, because institution 1, where both the quality and utility of feedback improved, had the highest participation rate in the discussion sessions. Future investigations should focus on tracking faculty self-study or ensuring that all faculty participate in the live discussion sessions.

3. Potential confounders: Some faculty may have personality traits that allow them to naturally be more comfortable with and able to provide better feedback than others. Also, some faculty have more experience with and/or may have had previous training in giving feedback, which may contribute to being able to give better feedback. Because we focused on a global assessment in this study, we did not collect demographic data on the faculty and residents or characteristic data on the institutions and cases for a multivariate analysis that would adjust for potential confounders. Future studies should collect data on and address these possible confounding variables because educating faculty may be only 1 factor contributing to the improvement of feedback quality and utility.
4. No control group: Because this study had a before-after design, we did not assess changes over time in a control group with faculty who were not exposed to the intervention. While the experience that faculty gained in providing feedback during the study period may have contributed to the changes we detected in this study, Baroffio et al²¹ found in a study on tutors

that the ability to provide feedback to students did not improve with experience.

5. Different feedback collection systems: Although all institutions collected daily open-ended feedback electronically, the systems they used were not exactly the same. To avoid confounders that may have arisen due to any system changes, we decided to utilize the systems that were previously established at each institution.

Despite these limitations, important strengths of this study were the blinded nature of the rating system and the multicenter design:

1. Blinded nature of rating system: In rating the feedback records, the raters were blinded to the data collection period, the institution, the faculty member providing the feedback, and the resident receiving the feedback. All feedback was rated by the same 3 trained experts working as a team to ensure quality and consistency of the ratings. This process avoided potential bias or issues with interrater reliability.
2. Multicenter design: Multiple centers were enrolled to ensure an adequate volume of data for analysis and to evaluate the generalizability of the educational program. The multicenter design proved critical in this study because, if the study had been undertaken only at the pilot center (institution 1), the educational program would have appeared successful for the 3 months following the intervention when it was not at other centers. This result suggests that more work needs to be done to make the intervention generalizable.

Other Training Programs

Since the initiation of this project, there has been some work in developing educational tools on providing feedback to anesthesiology residents. Dorotta et al¹⁰ provided a set of tools to teach and evaluate professionalism in residents. This system could facilitate and augment the tenets of feedback on professionalism taught in our sessions. In another study, Minehart et al⁹ demonstrated improvements in elements of feedback provision after a structured intervention, but this study was conducted in a simulated environment that may not reflect real-world conditions for providers or events.

Future Directions

While our feedback rating system did detect some changes, we plan to refine it based on our experience. Regardless of data collection period, the majority of feedback records were not destructive or harmful. This insight may suggest that the rating system should include other traits that can improve the discriminatory power of the system.

Our intervention was intentionally focused and much less elaborate than those previously described in the literature. We wanted to investigate whether a simple, easily reproducible intervention could impact feedback. While we detected changes at 1 institution, we seek to further explore other factors that may influence the impact of the intervention, including the time of year the feedback was provided, the level of training of the resident, and the faculty's experience with feedback. ■■

ACKNOWLEDGMENTS

The authors thank all the faculty and residents who participated in this study at Beth Israel Deaconess Medical Center (Boston, MA); University of California, San Diego (San Diego, CA); University of Kentucky (Lexington, KY); and University of Rochester Medical Center (Rochester, NY). They also thank Lisa Bartnikas, MD (Boston Children's Hospital, Boston, MA), Tom Bartnikas, MD, PhD (Brown University, Providence, RI), David Fobert, MLA (Beth Israel Deaconess Medical Center, Boston, MA), Erin Janssen, MD, PhD (Boston Children's Hospital, Boston, MA), Michael McBride, RN (Beth Israel Deaconess Medical Center, Boston, MA), Darren Tavernelli, RN, RRT (Beth Israel Deaconess Medical Center, Boston, MA), and H. Nicole Tran, MD (Kaiser Permanente Medical Care Program, Oakland, CA) for their assistance in producing the videos for the faculty development program; Lt Col Mary Jane Cahill, MBA, C-TAGME (Beth Israel Deaconess Medical Center, Boston, MA) for her work in the initial assessment of the feedback rating system; Ariel L. Mueller, MA (Beth Israel Deaconess Medical Center, Boston, MA) for her assistance in the statistical analysis; and the Center for Anesthesia Research Excellence (CARE, Beth Israel Deaconess Medical Center, Boston, MA) for their support of the study.

DISCLOSURES

Name: John D. Mitchell, MD.

Contribution: This author led the study design and implementation, data collection and analysis, and manuscript preparation. He also led the faculty development sessions at all the institutions and was one of the blinded experts rating the feedback records in the study.

Name: Cindy Ku, MD.

Contribution: This author helped with the data analysis as one of the blinded experts rating the feedback records in the study. She also helped with the interpretation of the data after analysis and reviewed manuscript drafts.

Name: Carol Ann B. Diachun, MD, MEd.

Contribution: This author helped with the study implementation and data collection at the University of Rochester Medical Center. She also helped with the interpretation of the data after analysis and provided guidance during the manuscript preparation phase.

Name: Amy DiLorenzo, MA.

Contribution: This author helped with the study implementation and data collection at the University of Kentucky. She also helped with the interpretation of the data after analysis and reviewed manuscript drafts.

Name: Daniel E. Lee, MD, PhD.

Contribution: This author helped with the study implementation and data collection at the University of California, San Diego. He also helped with the interpretation of the data after analysis and reviewed manuscript drafts.

Name: Suzanne Karan, MD.

Contribution: This author helped with the study implementation and data collection at the University of Rochester Medical Center. She also helped with the interpretation of the data after analysis and reviewed manuscript drafts.

Name: Vanessa Wong, BS.

Contribution: This author helped with the study design, data collection, data analysis, and manuscript preparation.

Name: Randall M. Schell, MD, MACM.

Contribution: This author helped with the study implementation and data collection at the University of Kentucky. He also helped with the interpretation of the data after analysis and reviewed manuscript drafts.

Name: Marek Brzezinski, MD, PhD.

Contribution: This author helped with the study design, assisted in the interpretation of the data after analysis, and provided guidance during the manuscript preparation phase.

Name: Stephanie B. Jones, MD.

Contribution: This author helped with the study design and implementation, data analysis, interpretation of the data after analysis,

and manuscript preparation. She led the resident sessions at all the institutions and was one of the blinded experts rating the feedback records in the study.

This manuscript was handled by: Edward C. Nemergut, MD.

REFERENCES

1. Ende J. Feedback in clinical medical education. *JAMA*. 1983;250:777–781.
2. Anesthesiology Milestone Group. *The Anesthesiology Milestone Project*. 2013. Available at: <http://acgme.org/acgmeweb/Portals/0/PDFs/Milestones/AnesthesiologyMilestones.pdf>. Accessed August 28, 2014.
3. Holmboe ES, Ward DS, Reznick RK, et al. Faculty development in assessment: the missing link in competency-based medical education. *Acad Med*. 2011;86:460–467.
4. Rosenblatt MA, Schartel SA. Evaluation, feedback, and remediation in anesthesiology residency training: a survey of 124 United States programs. *J Clin Anesth*. 1999;11:519–527.
5. Mitchell JD, Holak EJ, Tran HN, Muret-Wagstaff S, Jones SB, Brzezinski M. Are we closing the gap in faculty development needs for feedback training? *J Clin Anesth*. 2013;25:560–564.
6. Holmboe ES, Yepes M, Williams F, Huot SJ. Feedback and the mini clinical evaluation exercise. *J Gen Intern Med*. 2004;19:558–561.
7. Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Med Educ*. 2012;46:201–215.
8. Donato AA, Pangaro L, Smith C, et al. Evaluation of a novel assessment form for observing medical residents: a randomized, controlled trial. *Med Educ*. 2008;42:1234–1242.
9. Minehart RD, Rudolph J, Pian-Smith MC, Raemer DB. Improving faculty feedback to resident trainees during a simulated case: a randomized, controlled trial of an educational intervention. *Anesthesiology*. 2014;120:160–171.
10. Dorotta I, Staszak J, Takla A, Tetzlaff JE. Teaching and evaluating professionalism for anesthesiology residents. *J Clin Anesth*. 2006;18:148–160.
11. Salerno SM, Jackson JL, O'Malley PG. Interactive faculty development seminars improve the quality of written feedback in ambulatory teaching. *J Gen Intern Med*. 2003;18:831–834.
12. Mitchell JD, Ku C, Wong V, et al. The impact of a resident communication skills curriculum on patients' experiences of care. *A Case Rep*. 2016;6:65–75.
13. Van De Ridder JMM, Stokking KM, McGaghie WC, Cate Ten OTJ. What is feedback in clinical education? *Med Educ*. 2008;42:189–197.
14. Schlecht KD. Feedback. *Int Anesthesiol Clin*. 2008;46:67–84.
15. Cantillon P, Sargeant J. Giving feedback in clinical settings. *BMJ*. 2008;337:a1961.
16. Bhatnagar R, Kim J, E Many J. Candidate surveys on program evaluation: examining instrument reliability, validity and program effectiveness. *Am J Educ Res*. 2014;2:683–690.
17. Møystad A, Lycke KH, Barkvoll TA, Lauvås P. Faculty development for clinical teachers in dental education. *Eur J Dent Educ*. 2015;19:149–155.
18. Watling C, Driessen E, van der Vleuten CP, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Med Educ*. 2014;48:713–723.
19. Watling CJ. Unfulfilled promise, untapped potential: feedback at the crossroads. *Med Teach*. 2014;36:692–697.
20. Mitchell JD, Jones SB. Faculty development in feedback provision. *Int Anesthesiol Clin*. 2016;54:54–65.
21. Baroffio A, Kayser B, Vermeulen B, Jacquet J, Vu NV. Improvement of tutorial skills: an effect of workshops or experience? *Acad Med*. 1999;74:S75–S77.