

Improving Faculty Feedback to Resident Trainees during a Simulated Case

A Randomized, Controlled Trial of an Educational Intervention

Rebecca D. Minehart, M.D., Jenny Rudolph, Ph.D., May C. M. Pian-Smith, M.D., M.S., Daniel B. Raemer, Ph.D.

ABSTRACT

Background: Although feedback conversations are an essential component of learning, three challenges make them difficult: the fear that direct task feedback will harm the relationship with the learner, overcoming faculty cognitive biases that interfere with their eliciting the frames that drive trainees' performances, and time pressure. Decades of research on developmental conversations suggest solutions to these challenges: hold generous inferences about learners, subject one's own thinking to test by making it public, and inquire directly about learners' cognitive frames.

Methods: The authors conducted a randomized, controlled trial to determine whether a 1-h educational intervention for anesthesia faculty improved feedback quality in a simulated case. The primary outcome was an analysis of the feedback conversation between faculty and a simulated resident (actor) by using averages of six elements of a Behaviorally Anchored Rating Scale and an objective structured assessment of feedback. Seventy-one Harvard faculty anesthesiologists from five academic hospitals participated.

Results: The intervention group scored higher when averaging all ratings. Scores for individual elements showed that the intervention group performed better in maintaining a psychologically safe environment (4.3 ± 1.21 vs. 3.8 ± 1.16 ; $P = 0.001$), identifying and exploring performance gaps (4.1 ± 1.38 vs. 3.7 ± 1.34 ; $P = 0.048$), and they more frequently emphasized the professionalism error of failing to call for help over the clinical topic of anaphylaxis (66 vs. 41%; $P = 0.008$).

Conclusions: Quality of faculty feedback to a simulated resident was improved in the interventional group in a number of areas after a 1-h educational intervention, and this short intervention allowed a group of faculty to overcome enough discomfort in addressing a professionalism lapse to discuss it directly. (*ANESTHESIOLOGY* 2014; 120:160-71)

DIRECT and timely formative feedback to improve a learner's ongoing practice is one of the strongest predictors of improved performance in learning.¹⁻⁷ Such feedback conversations are difficult for faculty because of three challenges: (1) faculty worry that honest task-relevant critique will harm their relationship with the learner,⁸⁻¹¹ (2) overcoming cognitive biases impeding faculty's ability to diagnose the cognitive frames driving trainees' performances^{10,12-15} and tailoring the feedback conversation appropriately,¹⁶ and (3) working with time pressure that makes the first two challenges even more acute.

Faculty typically resolve the tension they experience between the "task *versus* relationship" dilemma in feedback conversations by emphasizing one over the other. Although skillful direct task feedback can be effective,^{12,17,18} harsh task feedback can: (1) degrade performance;¹⁷⁻¹⁹ (2) prevent reflection,¹³ absorption, and retention of knowledge;²⁰⁻²³ or (3) harm the pair's capacity to talk about difficult topics in the

What We Already Know about This Topic

- Feedback conversations are a critical part of effective teaching
- Overcoming concerns about the faculty-resident relationship and understanding the learner's frame of reference is challenging
- The investigators conducted a randomized trial evaluating the effect of a 1-h simulation-based training session on feedback quality

What This Article Tells Us That Is New

- Training improved faculty ability to maintain a psychologically safe environment, explore the resident's frame of reference, and address professionalism along with technical issues

future.^{24,25} Alternatively, feedback that emphasizes relationship preservation usually employs kindly leading questions and strategic silence to camouflage the instructor's critique and guide the learner to the instructor's hidden answer.^{8,24-26} Although benign in intent, this approach is time consuming, conveys the meta-message that errors are not discussible, and obscures important task-related feedback.^{24,26}

Submitted for publication May 15, 2013. Accepted for publication October 10, 2013. From the Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Boston, Massachusetts, and Center for Medical Simulation, Boston, Massachusetts.

Copyright © 2013, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins. *Anesthesiology* 2014; 120:160-71

Diagnosing learners' actual learning needs by uncovering the cognitive frames driving their actions poses a second challenge. Changing anyone's knowledge, skills, or attitudes requires learning by the teacher and the learner both at the level of visible actions and invisible cognitive processes.^{27–33} Faculty's ability to diagnose learner thought processes can be impaired by cognitive biases, such as mistaking their internal conclusions for reality,³⁴ erroneously estimating frequency of behaviors, or attribution error.^{35–38} These instructional mistakes are not the result of inadequate clinical expertise; rather, they are driven by ubiquitous cognitive biases propelling instructors to assume (often erroneously) that they know the reasons behind other people's actions.^{16,39}

Given the above conditions, specific, corrective feedback is exceedingly rare, measured as less than 3% of all utterances in some teaching encounters.⁴⁰ Feedback on the delicate topic of professionalism is even rarer.⁴¹

We know little about how anesthesia faculty manage these feedback challenges, typical patterns of how anesthesia faculty provide feedback, or the impact of efforts to enhance their skills. A recent review of faculty development interventions found no assessment of faculty's teaching or feedback skills in the acute care setting.⁴² A handful of studies outside the field of anesthesia examined the impact of interventions to improve faculty and resident feedback skills in the context of precepting. They reported only modest positive impact.^{40,41,43,44}

To ascertain how anesthesia faculty handle these feedback challenges with and without training, and to characterize anesthesia faculty's feedback patterns, we conducted a randomized, controlled trial to test the impact of an educational intervention as measured by rating feedback skills in a simulated case. The intervention addressed (1) how to manage the perceived task *versus* relationship dilemma by holding generous inferences about a learner; (2) how to provide direct, specific feedback pairing a statement of observable action ("advocacy") from the instructor's point of view, with an open-ended question ("inquiry") designed to determine the learners' thought processes;^{26,29,30,45} (3) how to address a specific lapse in professionalism by using a combination of these techniques.

Materials and Methods

With approval of the Institutional Review Board (Boston, MA), a balanced randomization (1:1), rater-blinded, parallel-group-controlled experiment was conducted during a recurring, mandatory, simulation-based crisis management course for practicing anesthesiologists from five academic hospitals in greater Boston, Massachusetts (Children's Hospital of Boston, Brigham and Women's Hospital, Massachusetts General Hospital, Beth Israel Deaconess Medical Center, and Newton-Wellesley Hospital). To detect an improvement in feedback of 2 of 3

of a unit of the rating scale with a one-sided 5% significance level and a power of 80%, a sample size of at least 30 subjects per group was required. It was anticipated that 10 subjects' cases would be needed for rater training and there would be other sources of attrition of subjects. Thus, the experiment was conducted during 71 regularly scheduled (approximately weekly) 6-h simulation-based teamwork and communication courses, each with four simulation scenarios from March 2008 to February 2010. In advance of the first course session, each course day was designated as an intervention day or a control day from a computer-generated random number table. Each course hosted between three and eight anesthesiologists. No more than two individuals from any of the five hospital departments were present during a given course. At the beginning of the course day, a single course participant was randomly selected to be the experimental subject by arbitrarily assigning an integer number to each participant and rolling one or two dice. Figure 1 shows a flow diagram of the experiment.⁴⁶

Each course day started with a 1-h standardized introduction to the course and a simulation scenario and debriefing unrelated to this experiment. After the unrelated scenario and debriefing, participants in the intervention group were exposed to the intervention itself, a 1-h didactic, video assessment, and role-play workshop on how to resolve the perceived task *versus* relationship dilemma, how to diagnose trainees' learning needs, and how to address different kinds of errors including professionalism lapses (appendix 1). The videos in this case were three prepared examples of a simulated faculty member engaged in a feedback conversation with a simulated resident after the resident committed a breach in sterility. Each video depicted different feedback skills and demonstrated the advantages of a "frame-based feedback" approach. Then the experimental case scenario was conducted. On control days, the 1-h workshop on principles of giving effective feedback was conducted after the experimental scenario was completed.

The experimental scenario was designed with two parts. The first part allowed the subject to observe a resident commit four errors while managing a simulated patient. In the second part, the subject conducted a feedback conversation with the resident about his performance. For each error, the resident had a series of standardized unspoken reasons driving his actions (known as frames). This frame remained hidden unless the subject made appropriate inquiries during the feedback portion of the case scenario. Table 1 lists the four errors and their associated frames.

Scenario

All participants were given a written clinical stem of the case before its starting. They were told that they would watch an anesthesia resident at the "end of his first year" (CA1, or PGY-2) give an anesthetic from an observation window outside the operating room. Then one of them would be asked

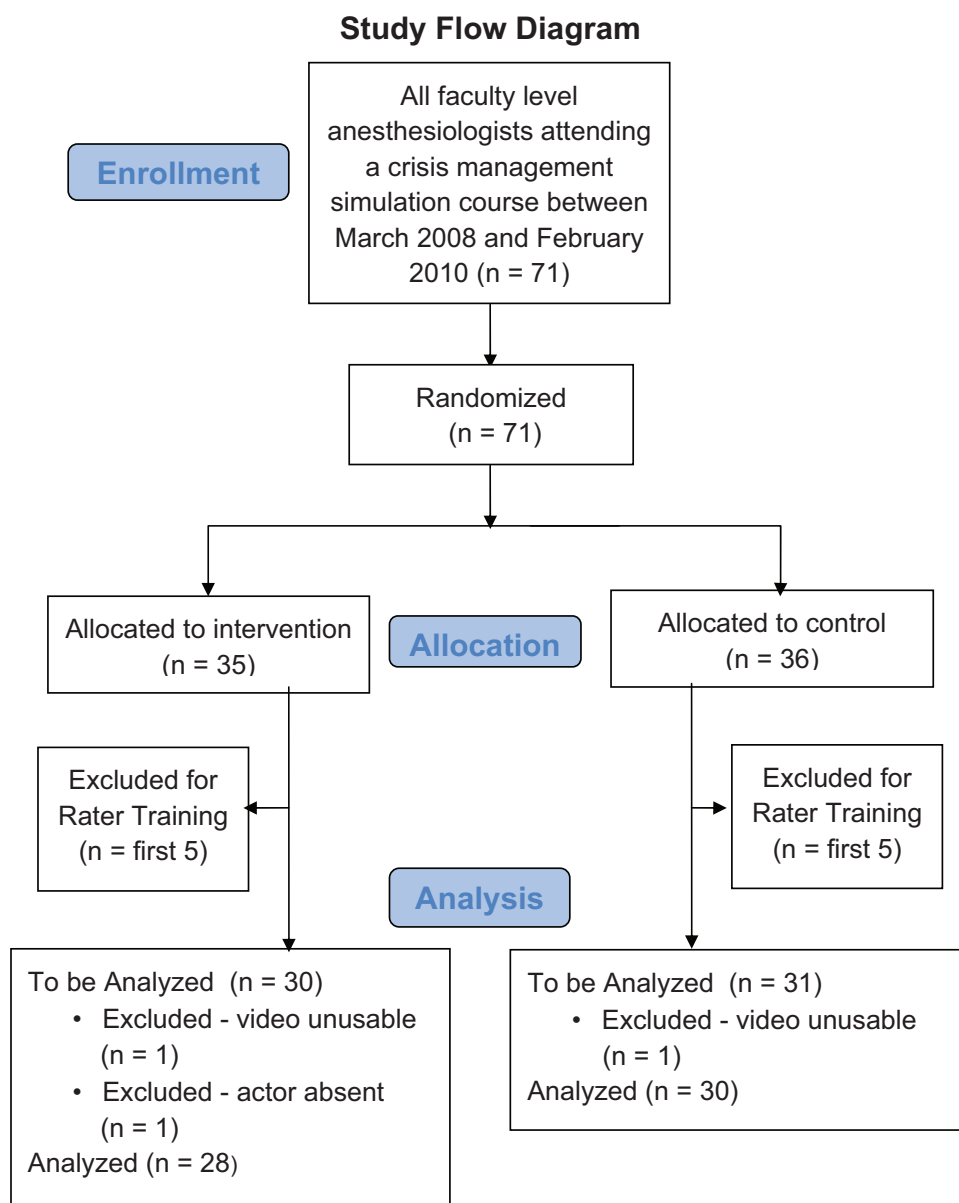


Fig. 1. Study flow diagram.

to give feedback to the resident on his conduct of the case. All participants, including the subject, were then brought to the observation window where they watched an operative procedure starting.

The patient was a 25-yr-old woman with a medical history significant only for hypertrophic obstructive cardiomyopathy who was undergoing an open appendectomy. There were four clinical actors in the scenario: a surgeon, scrub technician, circulating nurse, and an anesthesia resident ("resident"). The patient had already been induced under anesthesia and intubated, was stable with a heart rate (HR) of 60, blood pressure of 115/75 mmHg, and SpO₂ of 99%. During the surgeon's surgical prep, the patient's HR rose to 70 beats/min and the blood pressure fell to a systolic blood

pressure of approximately 95 mmHg. The surgeon noticed this and asked whether the patient was okay, to which the anesthesia resident responded that he would need to increase the anesthetic agent to blunt the rising HR. The anesthesia resident increased the sevoflurane-inhaled concentration from 1 to 4% (with high fresh gas flows), a relative overdose. The HR returned to 65 beats/min and systolic blood pressure to 100 mmHg, and the surgeon inquired whether the antibiotic had been given. This distracted the anesthesia resident from decreasing the sevoflurane, which led to commission of Error 1, leaving the sevoflurane at 4%.

The surgeon requested a 2-min warning to allow the antibiotic to circulate, which led to the resident giving the entire dose as an intravenous (IV) bolus. The HR rose

Table 1. Simulated Errors Committed by Junior Anesthesia Resident in Managing a Complex Case of a Patient with HOCM

Error Number	Type of Error from Bosk	Error Description	Resident's Frame
1	Technical	Resident increases vaporizer concentration to quickly deepen anesthesia, but forgets to turn it back down.	Has had a problem remembering steps that must be revisited. Has not been taught any "practice tricks" for this (such as leaving hand on vaporizer).
2	Judgmental	Resident focuses on regulating heart rate in patient with HOCM and misses onset of anaphylaxis to antibiotic.	Has never seen anaphylaxis and assumed it was really quite rare. Was focused on keeping the heart rate low to manage the heart rate-mediated fall in cardiac output seen in HOCM. Understands HOCM physiology and management quite well from preparatory reading. Also, knows how to treat anaphylaxis from lectures and reading.
3	Normative	Resident does not call for help as the patient's vital signs deteriorate and nurse asks repeatedly if help is wanted.	Has been directly criticized by this particular supervising attending for calling for help too early. Worries about getting a reputation as "a weak resident."
4	Quasinormative	Resident has not prepared a vasoactive medication for a patient with HOCM because he believes his supervisor is particular about not wasting drugs.	Has learned many of the practice preferences of the attending anesthesiologists he has worked with. This particular attending has strong opinions and an already low opinion of the resident's work.

HOCM = hypertrophic obstructive cardiomyopathy.

to 70 beats/min and the systolic blood pressure fell to less than 100 mmHg, but, unlike the previous episode, the SpO_2 decreased to 90%, the peak inspiratory pressure increased to 35 cm H_2O , and the ETCO_2 waveform developed a sloped early expiratory phase, indicative of respiratory obstruction (from anaphylaxis). The resident then gave esmolol (20 mg IV bolus) in response to the increased HR (commission of Error 2). The circulating nurse, noticing the deteriorating vital signs, asked whether the resident would like help, which the resident declined. After further deterioration of the vital signs (blood pressure = 80/60 mmHg, HR 90 beats/min, SpO_2 = 85%, peak inspiratory pressure = 40 cm H_2O , ETCO_2 = 28 mmHg with a sloped early expiration phase), the circulating nurse again asked whether she should call the attending anesthesiologist; the resident again declined help (commission of Error 3).

The surgeon then inquired about a rash that had just developed ("Does this patient have a rash?"), and the resident replied that the patient's skin was blotchy at the start ("I don't know, her skin was kind-of blotchy to begin with....").

At this point the simulation was paused. One of the investigators joined the participants at the now-closed observation window to conduct and guide a brief (3–5 min) discussion of the case to this point, to ensure a differential diagnosis was generated with anaphylaxis at the top of the list. Then one of the other participants (not the study subject) was asked to go into the operating room to help the resident, and the simulation was resumed.


The case continued with compromised vital signs remaining, and was then under the direction of the course participant. The resident was cooperative and competent. If asked to give phenylephrine, the resident would reply, "Ok, but I have to make up a drip [infusion]. My attending doesn't like me to make up phenylephrine ahead of time as he says it is a waste of money," (commission of

Error 4). If the participant did not ask for phenylephrine, the resident asked whether phenylephrine would be useful in this case, and then made the statement above regarding a previous attending's preference. The resident treated the patient in whatever manner directed by the course participant. If the initial treatment did not include epinephrine, the resident would suggest treating the anaphylaxis with a small dose of epinephrine in spite of the relative contraindication in a patient with hypertrophic obstructive cardiomyopathy. Once treated with epinephrine in any dose (all course participants gave epinephrine within a 5-min period), the patient recovered with near-normal vital signs and the simulation was ended.

Upon completion of the simulation case, the course participants returned to a debriefing room where a trained facilitator (D.B.R. or J.W.R.) conducted a discussion of four types of errors as described by Bosk.⁴⁷ Furthermore, four of the errors made by the observed resident were listed on a whiteboard and put into one of the four categories.

These errors were briefly discussed among participants and in all cases consensus was reached that the Normative Error of refusing to call for help was the most troubling.

The experimental subject was asked to move to "the coffee room" where the resident was taking a short break between cases, to give the resident feedback on his performance. The subject was suitably oriented to this exercise. Throughout the feedback session, the resident responded to the subject in a semiscripted manner according to a predetermined set of rules (appendix 2). The resident would try to keep the conversation as normal as possible, responding to direct questions only, answering to the ability of a competent resident nearing the completion of his CA1 year, and would be slightly evasive or defensive if asked vague questions about his performance. When offered an observation and asked relevant questions, the

 Video # _____ Scoring Date _____ Time _____ Rater Name _____

Feedback Assessment for Clinical Education (FACE)® Score Sheet

Rating	1	2	3	4	5	6	7
Descriptor	Extremely Ineffective	Consistently Ineffective	Mostly Ineffective	Somewhat Effective	Mostly Effective	Consistently Effective	Extremely Effective

Element 1 Establishes a safe feedback environment. <ul style="list-style-type: none"> Optimizes physical environment for conversation Builds rapport Establishes process expectations for the feedback session Articulates objectives for the feedback Commits to respecting recipient's understanding and perspective 	Element 1 Rating:
Element 2 Maintains a safe feedback environment. <ul style="list-style-type: none"> Maintains physical environment conducive for conversation Maintains rapport Follows or explicitly recrafts established process for the session Conveys a consistent commitment to respecting the recipient and understanding their perspective 	Element 2 Rating:
Element 3 Structures the feedback session in an organized way. <ul style="list-style-type: none"> Session progresses logically rather than jumping around Encourages recipient to express his/her reactions and, if needed, orients him/her to what happened in the clinical event Guides analysis of the recipient's performance Collaborates with the recipient to summarize learning from the session 	Element 3 Rating:
Element 4 Provokes engaging discussion. <ul style="list-style-type: none"> Uses concrete examples as the basis for discussion Reveals own reasoning and judgments Facilitates discussion through verbal and non-verbal techniques Recognizes and manages the upset recipient 	Element 4 Rating:
Element 5 Identifies and explores performance gaps. <ul style="list-style-type: none"> Provides feedback on performance Explores the source of the performance gap 	Element 5 Rating:
Element 6 Helps recipient achieve or sustain good future performance. <ul style="list-style-type: none"> Helps close the performance gap through effective teaching Demonstrates firm grasp of the subject Encourages reflection on the feedback Meets the important objectives of the session Assesses learning 	Element 6 Rating:

Fig. 2. Feedback assessment instrument using a six-element behaviorally anchored rating scale.

resident would reveal the first part of his mental frame, and would follow with more revealing of his mental frame if the subject pursued the topic further. If the subject lectured or taught about something that was not directly relevant to the resident's frame, the resident would listen and respond neutrally only with a nod or say "OK." If the subject seemed to be finished with the feedback session by explicitly saying so, by repeating the same material, or by summarizing the session, the resident would say that his pager had gone off and he needed to go do the next case.

After completing the feedback session, the subject returned to the debriefing room where an extensive discussion about the case and feedback was facilitated by one of the investigators (D.B.R. or J.W.R.).

The feedback session in the coffee room was video recorded for later analysis.

Feedback Performance Rating

The investigators developed a two-part rating instrument to assess the performance of the subjects in giving feedback. The first part was a behaviorally anchored rating scale (BARS) with six elements (fig. 2). This part was modified from another BARS tool that has been used for simulation debriefing, a specific form of feedback conversation, and has been partially validated.⁴⁸ The average rating of the elements was intended to be a measure of feedback performance and was the primary outcome variable for this study. The second part of the instrument was an objective scoring of 12 aspects of the feedback session designed to capture patterns of feedback (fig. 3).



Video # _____ Scoring Date _____ Time _____ Rater Name _____

Objective Rating

1. Discernible Reactions Phase

___(1) Yes (ex. how do you feel?) ___(2) No (jumps right into topic or how is patient?)

2. Signals that this will be a feedback session

___(1) Explicit statement ___(2) Vague statement or request (Ex. can we talk?) ___(3) None

 ___(1) Asks explicitly for **consent** for giving feedback or other mutual interaction?

3. Framing /roadmapping of the conversation? (Ex. I would like to talk about A and then B; topical)

___(1) One framing statement at beginning ___(2) Multiple instances of framing ___(3) No use of framing

4. Feedback giver's emphasis on the major topics (Not by time, but by focus)

Extensive	Some	Minimal	None	Issue
(1)	(2)	(3)	(4)	Physiology and management of HOCM
(1)	(2)	(3)	(4)	Missing diagnosis of anaphylaxis
(1)	(2)	(3)	(4)	Treatment of anaphylaxis
(1)	(2)	(3)	(4)	Not calling for help
(1)	(2)	(3)	(4)	Not having phenylephrine drip
(1)	(2)	(3)	(4)	Leaving Agent on
(1)	(2)	(3)	(4)	Other issue (specify)

5. Predominant style of feedback (check one only)

Advocating / Didactic / Lecturing	Advocating, but leading questions	Paired Advocacy and Inquiry	Open-ended questions	GWIT questions
(1)	(2)	(3)	(4)	(5)

6. Predominant conversational pattern of faculty

One-way talking	Balanced, but more talking	Balanced talking and listening	Balanced, but more listening	One-way listening
(1)	(2)	(3)	(4)	(5)

7. Describe your perception of the Faculty's attitudes (check one from each row)

 ___(1) **Activated** (engaged, interested) ___(2) **Neutral** ___(3) **Deactivated** (impatient, detached)

 ___(1) **Pleasant** (upbeat, relaxed) ___(2) **Neutral** ___(3) **Unpleasant** (angry, frustrated)

8. How might you have felt if you were in the learner's seat in this session? (check one from each row)

 ___(1) **Activated** (engaged, interested) ___(2) **Neutral** ___(3) **Deactivated** (disengaged, bored)

 ___(1) **Pleasant** (encouraged, respected, trusting) ___(2) **Neutral** ___(3) **Unpleasant** (shamed, interrogated)

9. Did they check the learner's understanding of the feedback?

___(1) Never ___(2) Occasionally (1-2 times) ___(3) Sometimes (3-4 times) ___(4) Frequently (>4 times)

10. Closure

___(1) None ___(2) Brief summary ___(3) Extensive (incl. action items) ___(4) invite F/U ___(5) N/A (ended)

11. Do you have the impression that the Faculty took the feedback simulation seriously?

___(1) Definitely not ___(2) Possibly not ___(3) Probably yes ___(4) Definitely yes (did something beyond)

12. Do you personally know the Faculty?

___(1) Definitely not ___(2) Casual acquaintance ___(3) Pretty well ___(4) Close friend

Fig. 3. Feedback assessment instrument using objective observation of 12 aspects of the feedback conversation. F/U = follow up; HOCM = hypertrophic obstructive cardiomyopathy.

This instrument was developed to mirror some of the generic feedback skills gleaned from the literature and expected to be discussed during the educational intervention.⁴⁹

Four blinded raters experienced in simulation, debriefing, and precepting residents or nurses, from different specialties (nursing education, endocrine surgery, pediatric anesthesiology, and intensive care medicine), were selected and trained to use the two-part instrument. Approximately 12 h of rater training was conducted using the first 10 videos, and those videos were then removed from the pool of study videos.⁵⁰ Adequate calibration of raters was confirmed upon completion of training, when all the raters were within one point of each other on the three test videos. See appendix 3 for a full description of the rater training process.

Of the remaining 61 videos, three were excluded from performance rating because in two cases the video recording was found to be unusable and in one case the resident actor was absent and a substitute was used. Thus, video files from 30 control and 28 intervention cases, all using the same resident actor, were randomly distributed to the four raters such that each rater had a balanced number of cases from each group and two raters would rate each video.

Statistical Analysis

The institution at which the participants practiced anesthesia was compared between the intervention and control groups using a chi-square test. In addition, the years since completing medical school and anesthesia residency were computed

Table 2. Demographics of Subjects Analyzed

Element	Intervention	Control	P Value
Number analyzed	28	30	
Years since medical degree granted (mean yr \pm SD)	22.5 \pm 9.5	19.7 \pm 9.5	0.26
Years since anesthesia residency completed (mean yr \pm SD)	14.8 \pm 10.1	13.6 \pm 10.2	0.72
Sex (%female:%male)	34:66	37:63	0.86
Institution (%A:%B:%C:%D:%E)	20:30:37:10:3	24:14:38:14:10	0.65

Institutions A, B, C, D, and E, represent the five Harvard-affiliated hospitals, but in a blinded fashion. They are (in alphabetical order, but not related to A–E designations above): Beth Israel Deaconess Medical Center, Brigham and Women's Hospital, Children's Hospital Boston, Massachusetts General Hospital, and Newton-Wellesley Hospital.

Table 3. Rating of Feedback Quality for Each Element of the BARS

Element	Intervention (I) or Control (C)	Ratings N	Mean Rating	SD of Ratings	P Value
Establishes a safe feedback environment	I	52	3.9	1.26	0.074
	C	64	3.6	1.13	
Maintains a safe feedback environment	I	64	4.3	1.21	0.004*
	C	65	3.8	1.16	
Structures the feedback session in an organized manner	I	64	4.2	1.19	0.013*
	C	65	3.8	1.10	
Provokes engaging discussion	I	64	4.3	1.29	0.061
	C	65	3.9	1.33	
Identifies and explores performance gaps	I	63	4.1	1.38	0.049*
	C	65	3.7	1.33	
Helps recipient achieve or sustain good future performance	I	63	4.2	1.38	0.081
	C	65	3.9	1.25	

* Denotes statistical significance ($P < 0.05$ used as limit).

BARS = Behaviorally Anchored Rating Scale; C = control group; I = intervention group.

for the participants in the intervention and control groups. A Mann–Whitney U test was used to test whether the experience level of the groups was different.

All ratings were entered into a spreadsheet for analysis. The consistency of feedback performance rating for all pairs of scores was analyzed by first averaging and rounding and then computing the frequency of scores within one point. Cohen's κ -statistic with quadratic rating was used on these values as another measure of interrater reliability (Vassar-Stats*). Videos with mean absolute differences in scoring greater than 1.5 points were independently rerated by one of the investigators blinded to the experimental conditions. A *tertium quid* model of score resolution was used where the two closest scores were included in further analyses.⁵⁰

All the ordinal performance ratings for the intervention and control group were computed and reported as mean \pm 1SD and compared using a Mann–Whitney U test (Vassar-Stats*), with P value less than 0.05 considered significant. Performance ratings for each of the six elements for the intervention and control group were similarly compared.

The raters scored the 21 objective measures for each feedback performance video and the results were transferred to a spreadsheet. The frequencies of the dichotomous scores of each objective measure were computed for the control, intervention,

and combined groupings. Where appropriate, a chi-square statistic was used to compare intervention to control groups, with P value less than 0.05 considered significant (Vassar-Stats*).

Results

Table 2 shows the demographic distribution of the subjects. There were no differences in institution or experience between the intervention and control group course participants.

Interrater Agreement

Two raters assessed each case except for one case that was assessed by all four raters due to an error in distributing the videos. Including the case assessed by the four raters, 66% of all scores and 81% of averaged element scores were within one point. Excluding the case assessed by the four raters, moderate agreement was seen ($\kappa = 0.53$). Sixteen cases revealed a between-rater average difference of 1.5 points or greater and these cases were rated by a third rater independently, who was one of the investigators who did not participate in the conduct of the experiment. Subsequent to score resolution, 88% of all score pairs were within one point and substantial agreement was achieved ($\kappa = 0.72$).

Quality of Feedback Measured by BARS

Averaging ratings across all six elements of the feedback rating scale, the intervention group scored higher (4.2 ± 1.28)

* Available at: <http://vassarstats.net>. Accessed September 20, 2012.

than the control group (3.8 ± 1.22 ; $P < 0.0001$). Scores for individual elements are shown in table 3.

General Patterns of Feedback Assessed by Objective Structured Instrument

Twenty-eight percent of subjects in both groups executed an “entry phase” to the conversation by giving the resident an opportunity to state his or her reactions to the scenario or how the patient fared. Subjects in both the intervention and control groups explicitly stated they were about to give feedback in 49% of the cases and explicitly asked permission to do so in 20%. We analyzed the frequency of faculty members “previewing,” or verbally signaling that a feedback conversation was about to occur (an example would be, “I’d like to talk with you about your performance just now and give some feedback on A, B, and C.”). Subjects in the intervention group (18%) were more likely to use a preview statement to outline the upcoming feedback than the control group (10%; $P = 0.05$). Subjects in the intervention group more commonly used advocacy/inquiry language (24 vs. 9%, respectively; $P = 0.04$) and less commonly used “guess what I am thinking” questioning (13 vs. 34%, respectively; $P = 0.01$) than their peers in the control group as a primary style of giving feedback.

Intervention and control group subjects did not differ in their balance of talking and listening, with 66% being rated as talking more than listening, 32% balancing talking and listening, and 2% predominantly listening.

Subjects were rated as seeming activated (79%), neutral (16%), deactivated (4%), and displaying a pleasant (72%), neutral (22%), and unpleasant (6%) attitude and were not different between groups. Raters scored their perception of how a trainee might feel with the given feedback as activated (68%), neutral (21%), and deactivated (11%); and pleasant (59%), neutral (21%), and unpleasant (20%) regardless of the subject grouping.

Thirty-five percent of subjects in both groups never checked the resident’s understanding of any feedback provided by the faculty. Closure, a conversation that consists of some form of summary or planning of next steps, was seen in 48% of the cases of feedback conversations in both the control and intervention groups.

The control and intervention groups differed in the emphasis they placed on different topics for feedback. Looking at how faculty distributed their time in the conversation, we found that the intervention group showed a greater frequency of emphasis on the *normative/professionalism error of failing to call for help or accepting help when offered* over the *anaphylaxis treatment algorithm* than the control group did (66 vs. 41%; $P = 0.008$). Conversely, emphasizing the *anaphylaxis treatment algorithm* over the *normative/professionalism error regarding help* was seen more commonly in the control group than in the intervention group (39 vs. 21%, respectively; $P = 0.04$).

Subjects in the intervention and control groups did not differ in some of the other topics such as *managing hypertrophic obstructive cardiomyopathy* (66%), *leaving the anesthetic agent at a high level in the face of hypotension* (43%), other topics (25%), and *not having a phenylephrine infusion prepared* (14%).

Discussion

We were able to demonstrate a significant difference in the quality of feedback (as defined by the average rating elements of the BARS tool) that faculty gave to a simulated resident committing a series of scripted errors after a short educational intervention. The intervention group received higher scores than the control group. The intervention group also scored higher on creating and maintaining a psychologically safe environment and exploring the simulated resident’s frame, and more frequently addressed the difficult-to-discuss professionalism lapse of rejecting offers to call for help.

Baseline patterns of feedback conversations were also established for the groups. The raters judged 59% of the feedback conversations as potentially causing the resident to feel activated and pleasant, whereas 20% were judged as potentially causing him to feel “unpleasant.” Our control and intervention groups of academic faculty members did not differ from each other on other confounding factors, and may therefore represent the population of faculty needing to provide feedback. Review of the feedback literature yielded many opinion articles and commentaries, and various theoretical methods have been published regarding how best to deliver or facilitate feedback. Theories and strategies are emerging about creating psychological safety, uncovering learners’ frames to tailor teaching, and holding generous inferences about learners to foster curiosity in the instructor. On the basis of these theories, we chose to teach advocacy/inquiry⁴⁵ over other models (e.g., Pendleton’s self-assessment or the feedback “sandwich”)⁴⁹ because it is a technique that efficiently solves a number of feedback challenges. (1) It allows faculty to quickly share and test the validity of their observations and opinion (the advocacy). (2) It ascertains the cognitive frames driving the learner’s performance or performance gaps (the inquiry). (3) By encouraging faculty to hold generous inferences about learners (e.g., the learner is capable, wants to learn, wants to improve) and curiosity about the learner’s thinking, it helps them resolve the difficult tension they feel regarding attending to task *versus* relationship. The generous inferences and curiosity solidify the relationship whereas the advocacy plus inquiry allows for direct task feedback.

Our current understanding of what constitutes quality feedback continues to expand. Creating a psychologically safe environment for conversation is increasingly recognized as the *sine qua non* of learning in groups and dyads.^{51–54} Psychological safety is a person’s assessment that the situation is safe for interpersonal risk-taking such as exposing one’s reasoning, asking for help, and speaking up. Creating a psychologically safe environment is a key factor in resolving the

perceived task *versus* relationship dilemma; faculty rightly worry that providing direct feedback will feel hurtful when learners do not feel psychologically safe. The fact that the intervention group rated higher in creating a psychologically safe environment and were able to provide direct feedback suggests that the intervention helped faculty resolve the tension they experienced regarding the task *versus* relationship dilemma such that they could address both simultaneously.

The intervention group's skillfulness at giving direct feedback and its ability to explore the cognitive frame driving the resident's performance gap were improved. By definition of our rating scale, the control group was in the "mostly ineffective" to "somewhat effective" range, whereas the intervention group was somewhere between "somewhat effective" and "mostly effective," a substantively and statistically significant difference with regard to quality. This suggests that pairing an advocacy that reveals the instructor's point of view with an open-ended inquiry that elicits the learner's thinking may protect faculty from early closure on an erroneous hypothesis about what the resident needs, and allow faculty to discover the resident's actual thinking process. Making one's reasoning public and inquiring into others' reasoning is a hallmark of "reflective practice," a discipline developed to help professionals assess and improve subordinate, peer, and one's own cognitive routines and emotional reactions.^{33,55–58} This finding indicates the positive potential for helping faculty to teach residents by using a simple conversational rubric.

The control group covered the topic of anaphylaxis treatment more extensively than the intervention group. However, the simulated resident in our scenario already understood the treatment algorithm for anaphylaxis; he did not have a gap in medical knowledge in this area, but rather he had difficulty in applying his knowledge to the actual clinical situation. With only limited time available for face-to-face teaching, the control group more frequently chose to spend this precious time on a comfortable, clinical topic of anaphylaxis treatment rather than spending needed time discussing the most important, but presumably more difficult-to-discuss lapse in the resident's performance: not calling for or accepting help when indicated.

We helped intervention subjects address the more difficult topic of not calling for help by introducing Charles Bosk's⁴⁷ landmark study completed in 1976 of how surgery faculty used residents' errors as occasions for learning and socialization. This study presaged the barriers faculty face in many contemporary residency programs to address lapses in professionalism.⁴⁷ In Bosk's study, faculty characterized technical and clinical judgment errors as "forgivable" whereas "normative" errors, or lapses in professionalism, were considered "unforgivable." When such errors make it difficult or impossible for faculty to hold the resident in high regard (to believe the best of a resident), this makes the feedback conversation much more difficult.

In our intervention, we contrasted the idea of "unforgivable errors" with the idea of holding "generous inferences"

about the resident, at least for the period of the conversation. By generous inferences we mean something like "the resident is intelligent, capable, trying to do the right thing," or "innocent until proven guilty," or "there is a 5% chance there is a good reason they did what they did." These inferences free the faculty member to overcome two barriers to good feedback: (1) it makes it easier to resolve the perceived task *versus* relationship dilemma because they are not having to cover up a negative assessment of the learner's character; and (2) it frees faculty to use their curiosity and diagnostic skills to understand the resident's perspective and learning needs and thereby better close performance gaps.

Our BARS was based on the Debriefing Assessment for Simulation in Healthcare,⁴⁸ and feedback is a form of debriefing conversation. Drawing from our understanding of these debriefing conversations in simulation, we evaluated specific techniques such as using a reactions phase (to allow the resident to "vent" his/her feelings, or as an icebreaker/conversation starter), previewing (signaling to the resident that a feedback conversation was to begin), and reaching closure (signaling the end of the conversation, summarizing). Reaching closure was only seen in half of the cases, and previewing or conducting a reactions phase were seen even more infrequently, which suggests that faculty may need additional time and effort to learn and practice these techniques.

Although the majority of feedback conversations were judged to make the resident feel activated and pleasant, one of five feedback conversations were judged to make the resident feel unpleasant. We do not know whether the first finding is due to emphasis on relationship at the expense of accurate task feedback or whether it is skillful feedback that was able to attend to both task and relationship. The percentage of interventions judged by raters as likely to elicit unpleasant reactions in the resident is quite large when considering the importance of this feedback session, and deserves more attention. Although we did not collect data on specifics of why raters felt the trainee would react negatively, this is an area of focus for the future, because it is likely that both language choice and nonverbal cues could have contributed.

There are a number of limitations of this study. First, the study was not designed to evaluate retention of feedback skills, and perhaps our findings for the intervention group could be explained by a "recency effect" whereby the recently provided instruction was easily accessed at the moment, without long-term effect. In addition, not all clinical conditions necessitating feedback were replicated, which could limit applicability of our findings. Second, as in any simulation, there is a risk that participants might not have taken the conditions seriously; however, in exit surveys done after each course, there was no indication that this was a problem for any participant. Furthermore, even if participants took the simulation seriously, they did know it was a simulation and this may have affected their performance in some way. Third, although the performance-rating system we used was based on a partially validated subjective scoring instrument used for debriefing,

it has not been used previously for clinical feedback conversations. Thus, its psychometric properties for this purpose are not well known. Fourth, there are a variety of alternative feedback models in the published literature. Further studies of these models could be helpful to compare and gain insight into whether a “best fit” for constructing feedback exists. Finally, although the raters were blinded, the resident actor was not. Even though he was working from a script, there may have been subtle differences in the resident’s responses to the feedback provided and this could have biased the results.

Although review of the literature on feedback has yielded many opinion articles and commentaries and a handful of qualitative empirical studies, we present this report of a randomized controlled trial of teaching feedback, which demonstrated a real and immediate effect. Most importantly, this short educational intervention allowed a group of faculty to overcome enough of the discomfort of addressing a professionalism lapse to discuss it directly. Further studies will be needed to address timing of faculty development sessions, need for upkeep of skills and duration of retention, and transferability to other practices and disciplines.

Acknowledgments

The authors thank the significant contributions of the faculty and staff at the Center for Medical Simulation, Boston, Massachusetts; the support of the Foundation for Anesthesia Education and Research, Rochester, Minnesota, which made possible this work; and Roy Phitayakorn, M.D., M.H.P.E., Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts, Laura K. Rock, M.D., Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, Massachusetts, Andres T. Navedo-Rivera, M.D., Harvard Medical School, Children’s Hospital Boston, Boston, Massachusetts, and Amy Smith, D.N.P., M.S.N., C.N.M., Clinical Learning and Simulation Centers, Boston College, William F. Connell School of Nursing, Chestnut Hill, Massachusetts.

This work was funded by a Foundation for Anesthesia Education and Research (FAER) Research in Education Grant (REG): FAER REG 08/15/2008 Minehart.

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. Minehart: Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, GRJ 440, 55 Fruit Street, Boston, Massachusetts 02114. rminehart@partners.org. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. *ANESTHESIOLOGY*’s articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

1. Harlen W, James M: Assessment and learning: Differences and relationships between formative and summative assessment. *Assess Educ Princ Pol Pract* 1997; 4:365–77
2. Hattie J, Jaeger R: Assessment and classroom learning: A deductive approach. *Assess Educ Princ Pol Pract* 1998; 5:111–25
3. Natriello G: The impact of evaluation processes on students. *Educ Psychol* 1980; 22:155–75
4. Rudolph JW, Simon R, Raemer DB, Eppich W: Debriefing as formative assessment: Closing performance gaps in medical education. *Acad Emerg Med* 2008; 15:1110–6
5. Swanson DB, Norman GR, Linn RL: Performance-based assessment: Lessons from the health professions. *Educational Res* 1995; 24:5–11
6. Torrance H, Pryor J: Defining and Investigating Formative Assessment, *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*. Florence, Taylor & Francis, Inc., 1998, pp 8–20
7. Torrance H, Pryor J: Formative Assessment and Learning: Where Psychological Theory Meets Educational Practice, *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*. Florence, Taylor & Francis, Inc., 1998, pp 83–105
8. Ende J, Pomerantz A, Erickson F: Preceptors’ strategies for correcting residents in an ambulatory care medicine setting: A qualitative analysis. *Acad Med* 1995; 70:224–9
9. Mann K, van der Vleuten C, Eva K, Armson H, Chesluk B, Dornan T, Holmboe E, Lockyer J, Loney E, Sargeant J: Tensions in informed self-assessment: How the desire for feedback and reticence to collect and use it can conflict. *Acad Med* 2011; 86:1120–7
10. Rudolph JW, Foldy EG, Robinson T, Kendall S, Taylor SS, Simon R: Helping Without harming: The instructor’s feedback dilemma in debriefing—A case study. *Simul Healthc* 2013; 8:304–16
11. Sargeant J, Armson H, Chesluk B, Dornan T, Eva K, Holmboe E, Lockyer J, Loney E, Mann K, van der Vleuten C: The processes and dimensions of informed self-assessment: A conceptual model. *Acad Med* 2010; 85:1212–20
12. Archer JC: State of the science in health professional education: Effective feedback. *Med Educ* 2010; 44:101–8
13. McIlwrick J, Nair B, Montgomery G: “How am I doing?”: Many problems but few solutions related to feedback delivery in undergraduate psychiatry education. *Acad Psychiatry* 2006; 30:130–5
14. Kunda Z: Heuristics: Rules of Thumb for Reasoning, *Social Cognition: Making Sense of People*. Cambridge, MIT Press, 1999, pp 53–110
15. Kunda Z: Hot Cognition: The Impact of Motivation and Affect on Judgment Motivation, *Social Cognition: Making Sense of People*. Cambridge, MIT Press, 1999, pp 211–64
16. Shute VJ: Focus on formative feedback. *Rev Educ Res* 2008; 78:153–89
17. Hunt EA, Fiedor-Hamilton M, Eppich WJ: Resuscitation education: Narrowing the gap between evidence-based resuscitation guidelines and performance using best educational practices. *Pediatr Clin North Am* 2008; 55:1025–50, xii
18. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB: Systematic review of the literature on assessment, feedback and physicians’ clinical performance: BEME Guide No. 7. *Med Teach* 2006; 28:117–28
19. Baron RA: Negative effects of destructive criticism: Impact on conflict, self-efficacy, and task performance. *J Appl Psychol* 1988; 73:199–207
20. Zhao N: Learning from errors: The role of context, emotion, and personality. *J Organ Behav* 2011; 32:435–63
21. Cederblom D: The performance appraisal interview: A review, implications, and suggestions. *Acad Manage Rev* 1982; 7:219–27
22. Weisinger H: Protect the Self Esteem, The Power of Positive Criticism. New York, AMACOM, 2000, pp 17–20
23. Weisinger H: Put Motivations in Your Criticisms, The Power of Positive Criticism. New York, AMACOM, 2000, pp 65–9
24. Kegan R, Lahey L: From the Language of Constructive Criticism to the Language of Deconstructive Criticism,

- How the Way We Talk Can Change The Way We Work. San Francisco, Jossey-Bass, 2002, pp 91–102
25. Stone D, Patton B, Heen S: *Difficult Conversations*. New York, Penguin Books, 1999
 26. Rudolph JW, Simon R, Rivard P, Dufresne RL, Raemer DB: Debriefing with good judgment: Combining rigorous feedback with genuine inquiry. *Anesthesiol Clin* 2007; 25:361–76
 27. Argyris C: *Transition to Model I to Model II, Intervention Theory and Method: A Behavioral Science View*. Reading, Addison-Wesley, 1970, pp 96–109
 28. Argyris C: *Learning Model II Behavior, Intervention Theory and Method: A Behavioral Science View*. Reading, Addison-Wesley, 1970, pp 110–38
 29. Argyris C, Putnam R, Smith DM: *Engaging the Learning Process, Action Science: Concepts, Methods and Skills for Research and Intervention*. San Francisco, Jossey-Bass, 1985, pp 276–318
 30. Argyris C, Putnam R, Smith DM: *Promoting Reflecting and Experimentation, Action Science: Concepts, Methods, and Skills for Research and Intervention*. San Francisco, Jossey-Bass, 1985, pp 319–67
 31. Bandura A: Social cognitive theory of self-regulation. *Organ Behav Hum* 1991; 50:248–87
 32. Carver CS, Scheier MF: *Discrepancy Reducing Feedback Processes in Behavior, On the Self-Regulation of Behavior*. Cambridge, England, Cambridge University Press, 1998, pp 29–47
 33. Schön D: The dialogue between coach and student, *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco, Jossey-Bass, 1987, pp 100–18
 34. Rudolph JW, Morrison JB, Carroll JS: The dynamics of action-oriented problem-solving: Linking interpretation and choice. *Acad Manage Rev* 2009; 34:733–56
 35. Bazerman MH, Moore D: *Introduction to Managerial Decision Making, Judgment in Managerial Decision Making*, 7th edition. New York, John Wiley and Sons, 2008, pp 1–12
 36. Bazerman MH, Moore D: *Common Biases, Judgment in Managerial Decision Making*, 7th edition. New York, John Wiley and Sons, 2008, pp 13–41
 37. Bazerman MH, Moore D: *Bounded Awareness, Judgment in Managerial Decision Making*, 7th edition. New York, John Wiley and Sons, 2008, pp 42–61
 38. Ross L, Anderson CA: Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments, *Judgment Under Uncertainty*. Edited by Kahneman D, Slovic P, Tversky A. Cambridge, Cambridge University Press, 1982, pp 129–52
 39. Rudolph J, Raemer D, Shapiro J: We know what they did wrong, but not why: The case for 'frame-based' feedback. *Clin Teach* 2013; 10:186–9
 40. Salerno SM, O'Malley PG, Pangaro LN, Wheeler GA, Moores LK, Jackson JL: Faculty development seminars based on the one-minute preceptor improve feedback in the ambulatory setting. *J Gen Intern Med* 2002; 17:779–87
 41. 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; 18:831–4
 42. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D: A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach* 2006; 28:497–526
 43. Gelula MH, Yudkowsky R: Microteaching and standardized students support faculty development for clinical teaching. *Acad Med* 2002; 77:941
 44. Zabar S, Hanley K, Stevens DL, Kalet A, Schwartz MD, Pearlman E, Brenner J, Kachur EK, Lipkin M: Measuring the competence of residents as teachers. *J Gen Intern Med* 2004; 19(5 Pt 2):530–3
 45. Minehart RD, Pian-Smith MC, Walzer TB, Gardner R, Rudolph JW, Simon R, Raemer DB: Speaking across the drapes: Communication strategies of anesthesiologists and obstetricians during a simulated maternal crisis. *Simul Healthc* 2012; 7:166–70
 46. Schulz KF, Altman DG, Moher D; CONSORT Group: CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010; 63:834–40
 47. Bosk CL: *Error Rank and Responsibility, Forgive and Remember: Managing Medical Failure*, 2nd edition. Chicago, United Kingdom, University of Chicago Press, 2003, pp 36–70
 48. Brett-Fleegler M, Rudolph J, Eppich W, Monuteaux M, Fleegler E, Cheng A, Simon R: Debriefing assessment for simulation in healthcare: Development and psychometric properties. *Simul Healthc* 2012; 7:288–94
 49. Cantillon P, Sargeant J: Giving feedback in clinical settings. *BMJ* 2008; 337:a1961
 50. Johnson RL, Penny JA, Gordon B: *Training Raters and Staff, Assessing Performance: Designing, Scoring, and Validating Performance Tasks*. New York, The Guilford Press, 2009, pp 190–24
 51. Edmondson A: Psychological safety and learning behavior in work teams. *Admin Sci Quart* 1999; 44:350–83
 52. Edmondson AC: Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *J Manage Stud* 2003; 40:1419–52
 53. Edmondson AE: Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *J Appl Behav Sci* 1996; 32:5–28
 54. Tucker A, Edmondson A: Why hospitals don't learn from failures: Organizational and psychological dynamics that inhibit system change. *Calif Manage Rev* 2003; 45:55–72
 55. Freshwater D: Reflexivity and intersubjectivity in clinical supervision: On the value of not-knowing, *Transforming Nursing through Reflective Practice*. Edited by Johns C, Freshwater D. Oxford, United Kingdom, Blackwell, 2005, pp 99–113
 56. Mann K, Gordon J, MacLeod A: Reflection and reflective practice in health professions education: A systematic review. *Adv Health Sci Educ Theory Pract* 2009; 14:595–621
 57. Raelin J: Public reflection as the basis of learning. *Manage Learn* 2001; 32:11–30
 58. Taylor S, Rudolph J, Foldy E: Teaching reflective practice in the Action Science/Action Inquiry tradition: Key concepts and practices, *Handbook of Action Research*, 2nd edition. Edited by Reason P, Bradbury H. Thousand Oaks, Sage, 2008, pp 656–8

Appendix 1. Educational Intervention

One-hour Workshop on Giving Effective Feedback

Learning goals:

- Appreciate the rationale for diagnosing the trainee's learning needs by using frame-based feedback;
- Appreciate the role of holding generous inferences about the learner as a way to create psychological safety and overcome the task *versus* relationship dilemma;
- Gain skill in pairing advocacy with inquiry to give direct, corrective feedback and elicit learner's frames to determine their learning needs.

Curriculum

- Interactive slide and video presentation and exercises
 - Cognitive frame model of learning lecture (5 min)
 - Analyze example of clinical error by using cognitive frame model (5 min)
 - Fundamentals of feedback; especially the role of assuming the best about learners as a starting point lecture (5 min)
 - Assess video examples of attending giving feedback to residents (10 min)
 - Feedback “algorithm” lecture; how to use previewing statement and pair advocacy (instructor’s opinion about performance paired with inquiry about learner’s frame) (5 min)
 - Coached role-play exercise by using algorithm (15 min)
 - Bosk error taxonomy lecture (5 min)
 - Categorize resident’s errors by using Bosk taxonomy (5 min)
 - Prepare participants to address not-calling-for-help error (5 min)

Appendix 2. Detailed Instructions for Both Study Subjects and Resident Actor

Instructions to Subject

She or he was also told that for the purpose of this exercise, she or he could assume that she or he had perfect knowledge of the case as an observer and that the resident would not find that surprising. On arrival at the simulated coffee room, the subject found the resident alone at a conference table using a computer. The subject was free to sit anywhere, shut or leave the door open, and attain any posture she or he chose. Upon engaging the resident, she or he was free to follow any lines of conversation.

Semiscripted Responses for Resident Actor

(1) While trying to keep the conversation as normal as possible, the resident responded to direct questions only and did not elaborate unless asked. (2) If the resident was asked questions to elucidate his knowledge of a clinical subject, he responded with correct answers at depth of knowledge expected from a competent trainee at the end of first year level (in the opinion of the investigators). For example, if asked about the physiology of hypertrophic obstructive cardiomyopathy, he would describe a cardiomyopathy whereby the outflow track of the left ventricle would become

obstructed if the heart rate were increased. (3) If asked questions about his performance, the resident replied evasively and defensively. For example, if asked whether he thought he asked for help early enough, he would reply that he was glad that someone came when he needed help. (4) If the subject made a clear observation and asked a relevant question, the resident revealed the first part of his mental frame. For example, if the subject said, “I noticed the circulating nurse asked you twice if you needed help and you declined. How come?” the resident would reply, “I wanted to be sure if the esmolol worked and that I had done everything I was supposed to do in that situation.” (5) If the subject pursued the topic with further questioning the resident would reveal the second part of his mental frame. For example, if the subject said, “It seemed that the esmolol was not helping the blood pressure and I thought you could have used the help. How did you see it?” the resident would reply, “Actually, I have been chastised by my attending before for calling for help prematurely. I think I have gotten a reputation for being weak and I know that is hard to shake.” (6) If the subject lectured or taught about something that was not directly relevant to the resident’s frame, the resident would listen and respond neutrally only with a nod or say “ok.” (7) If the subject seemed to be finished with the feedback session by explicitly saying so or by repeating the same material, or by summarizing the session the resident would say his pager had gone off and he needed to go do the next case.

Appendix 3. Rater Training Process

1. Raters read and discussed case materials.
2. An investigator introduced the instruments and discussed each item.
3. The raters watched as a group and independently rated a video.
4. An investigator and the raters discussed score matches and differences.
5. Items 3 and 4, above, were repeated until three videos were analyzed.
6. Each rater independently watched and rated three videos and the results were compared to see whether all element ratings were within one unit and objective measures matched.
7. Raters and an investigator repeated 3 and 4, above, for one more video.
8. Item 6, above, was repeated.