

Moderator: Heather Harrell, MD  
 Discussant: Rhee Fincher, MD

## Failure to Fail: The Perspectives of Clinical Supervisors

Nancy L. Dudek, Meredith B. Marks, and Glenn Regehr

### Abstract

#### Background

Clinical supervisors often do not fail students and residents even though they have judged their performance to be unsatisfactory. This study explored the factors identified by supervisors that affect their willingness to report poor clinical performance when completing In-Training Evaluation Reports (ITERS).

#### Method

Semistructured interviews with 21 clinical supervisors at the University of Ottawa

were conducted and qualitatively analyzed.

#### Results

Participants identified four major areas of the evaluation process that act as barriers to reporting a trainee who has performed poorly: (1) lack of documentation, (2) lack of knowledge of what to specifically document, (3) anticipating an appeal process and (4) lack of remediation options.

#### Conclusions

The study provides insight as to why supervisors fail to fail the poorly performing student and resident. It also offers suggestions of how to support supervisors, increasing the likelihood that they will provide a valid ITER when faced with an underachieving trainee.

*Acad Med.* 2005;80(10 suppl):S84–S87.

In-training evaluation for medical students and residents relies heavily on faculty ratings of their clinical performances. A study of American medical schools from 1992–1998 found that these ratings accounted for 50–70% of a student's summative clerkship grade.<sup>1</sup> Although the reliability and validity of this form of evaluation has been questioned, the heavy reliance on faculty ratings continues because there is often no alternative.<sup>2</sup> As a result, research has been directed toward improving these ratings.

Problems in performance appraisal are usually attributed to a weakness of the instrument or a lack of knowledge of how to use the instrument. A significant amount of research has examined these issues.<sup>2</sup> However, the overall quality of these evaluations has not fundamentally improved. One reason may be that there has been little consideration of the personal attributes of the rater and systemic factors that might influence how raters report a specific performance.<sup>2</sup>

It is generally believed that experts know what competent clinical performance

entails and can judge both the quality and appropriateness of the student's practice.<sup>3</sup> However, there is evidence to suggest that the final evaluation is not always consistent with the evaluator's judgment of the performance.<sup>4–6</sup> Many clinical educators agree that there is a problem with reporting unsatisfactory performance of medical students and residents.<sup>4,5,7,8</sup> Clinical teachers at McMaster University rated "unwillingness to record negative evaluations" as the single most important problem with evaluation.<sup>7</sup> The Association of American Medical Colleges (AAMC) surveyed clinical faculty at ten schools and "unwillingness to record negative evaluations" was rated as a problem by 74.5%.<sup>8</sup>

Many authors have speculated about possible reasons for the failure to report negative performances, such as an unwillingness to invest the time required if their evaluation is appealed, a lack of confidence in their assessment ability and a fear of legal repercussions.<sup>4,7,9</sup> However, little empirical research has been conducted to explore these possibilities.

The objectives of this study were to: (1) determine why supervisors do not report poor clinical performance when completing evaluations, and (2) investigate what interventions supervisors

feel might make reporting poor performance easier.

#### Method

Physicians in the Departments of Medicine and Surgery at the University of Ottawa involved in evaluating medical students and/or residents during clinical rotations were recruited using a purposeful sampling technique.<sup>10</sup> Physicians who had not supervised a student or resident within the previous year were excluded. Initial interviews were requested with residency program directors and clinical rotation coordinators. Those interviewed were asked to share the names of colleagues who might be able to provide additional insight. Care was taken to ensure the physicians interviewed represented various subspecialties, had a variable level of involvement in the education program, and included both junior and senior faculty members. Informed consent was obtained prior to the interview. The Ottawa Hospital Ethics Board approved this project.

Consistent with techniques of theoretical sampling, no a priori estimate of sample size was calculated.<sup>10</sup> Rather, analysis was performed in parallel with data collection until it was determined that saturation of the recurrent themes had been achieved.

Correspondence: Nancy Dudek, MD, 505 Smyth Road, Room 1105D, Ottawa, Ontario, Canada, K1H 8M2; e-mail: (ndudek@ottawahospital.on.ca).

Clinical supervisors were invited to participate in 45-minute, semistructured interviews designed to explore their perspectives on the evaluation of poorly performing students or residents. Interviews were audiotaped and transcribed verbatim. Identifying information was removed from the transcripts prior to analysis. Data were extracted using the process of open coding to identify recurring themes and associations by two researchers.<sup>10</sup> Data were reviewed and compared following every four interviews. Explanations and conclusions regarding the outcomes began to take form, using data collected from the initial interviews. Consistent with grounded theory, these explanations were repeatedly reviewed and refined as additional data were collected.<sup>10</sup> Particular care was taken to explore contradictory or negative cases, ensuring a wide range of perspectives were included.<sup>11</sup> Discussion between the two researchers served to resolve discrepancies. Consensus was reached and a confirmed coding structure was generated. This structure was used to code the entire data set using qualitative data analysis software (NUD\*IST – version N6).

## Results

Twenty-one participants were interviewed: 16 were male and five were female. Thirteen participants were from the Department of Medicine and eight were from the Department of Surgery. Twelve clinical supervisors had greater than 10 years experience while nine had less. No consistent differences were found when the data were compared across these dimensions, so the data were treated as a single set.

Five participants had never given a failing evaluation to a resident. One of these participants had overturned a failing evaluation because he had been threatened with legal action. Based on that experience, the respondent would not fail a student in the future unless there was gross professional misconduct. Another participant extended the rotation of a failing resident, allowing time for remediation. At the end of the extended rotation, the resident was performing within expected standards and given a passing grade. The other three participants who had not given a

failing evaluation felt that they had never worked with a failing trainee.

By contrast, only six of the 21 participants could recall failing a student. Five of these failures were in general surgery or medicine. Most of the subspecialists felt that they had not failed a medical student because they usually only worked with students doing elective rotations in which they had much lower expectations for the students.

Most participants felt confident in their ability to determine whether a trainee was performing poorly. This confidence increased with experience.

Once you've been on the service for seven years on and off, you become more competent in your abilities to recognize a failing trainee.

All of the participants commented that the type of evaluation form used is not that important. There was strong agreement that the written comments were more useful than the global rating scales or checklists.

Participants identified a sense of responsibility as the main motive to fail a trainee: to the public to ensure safety, to the profession to protect its reputation, and to the trainee to allow them the opportunity for remediation.

A lot of us have that feeling . . . Would we let this person operate on our father?

These are the people that are going to be your colleagues. Your reputation as a profession depends on it (a valid evaluation process).

We do see some residents come through who ultimately fail the Royal College exams. . . . And I always think, 'Why did we not do something before this person wasted so many years of their life?'

Participants identified four major areas as barriers to failing trainees: (1) lack of documentation, (2) lack of knowledge of what to specifically document, (3) anticipating an appeal process, and (4) lack of remediation options.

The most commonly reported barrier to failing a trainee was that the participants had not kept a record of the trainee's day-to-day performance. Therefore, when it came time to fill out the In-training Evaluation Report (ITER), supervisors lacked enough supporting evidence for their judgment. Some participants did

record more detailed information. However, some felt that the time it took to do this properly was somewhat onerous.

Often we don't do a good enough job of recording performance. Then if the student challenges you, you do not have a leg to stand on because you cannot recall specific incidents.

I didn't have as good documentation as I should have. I had to go back and retrospectively create that documentation because I didn't realize that that was an issue.

As a second, related barrier to failing trainees, participants also reported that they did not know what type of information should be documented to support their impression that the trainee was performing poorly. Some indicated that if they had been better informed, they would have kept appropriate documentation. Others did not know how to identify the specific behaviors that resulted in their impression that the student was failing.

While it's hard to translate it to paper, I think if you work with somebody for a bunch of weeks, you know whether they're a good doctor or not. The problem is, before you commit to paper, stating that they aren't good, you need something concrete, and often it's not concrete.

The formal appeal process was the third major barrier identified by many participants. Concerns regarding the appeals process arose in two very different forms. Those who had not been through an appeal believed that it would be time-consuming; those who had, acknowledged that fact.

The resident went through all five levels of appeal, which took a couple of years . . . a tremendous amount of work. All five levels upheld our decision, which is nice but the experience was awful.

Additionally, participants felt that the appeal process puts their credibility on the line.

It's just a pain . . . way more work, way more documentation. . . . You're going to have to defend your actions with the program director or at the university level.

As a result, despite their previously stated confidence in their ability to recognize a failing trainee, supervisors often look for confirmation from their colleagues that

the trainee had performed poorly on previous rotations. If they are unable to find support for their judgment, they are less likely to fail the trainee.

If I have an R3 and they're really stinking up the joint and failing the rotation, and I heard through the grapevine that maybe it had been an issue or there was some borderline performance before, I'd probably stick to my guns. If I went around and I heard across the board that this was a stellar resident, I would be very nervous about failing that resident.

Only one supervisor admitted that having been threatened in the past with legal action now prevents them from failing trainees. Although many supervisors admitted to being threatened, they felt that the time involved in an appeal was a greater barrier. One supervisor recalled changing a student's grade and now regrets it.

I felt afterwards that it was unfair to all the other students and it undermined my credibility.

Finally, many participants felt that they could not fail a trainee if remediation was not available to them. They felt it was their responsibility to provide remediation and, if unable to do so, they might not fail that person.

You fail somebody and then you go, well now what? Do we actually have a mechanism to help you get better? It (the lack of proper remediation) causes me to think twice before I fail somebody.

The participants were asked to identify potential solutions for making the process of reporting a failing trainee easier for them. Participants felt that they needed to learn how to document trainee performance in an efficient manner. Some felt that technology could be used to make recording information on the daily performance of trainees easier.

People nowadays are very comfortable typing in comments and filling out forms electronically.

The participants also said they need to know what type of evidence needs to be documented. There was strong support for faculty development workshops, where supervisors could learn how to implement the evaluation process on a day-to-day basis. This was felt to be particularly important for new staff.

We have faculty development on how to teach and how to mentor . . . but we don't

have faculty development on evaluation . . . It should be just as important as teaching a resident to do a complicated procedure.

Finally, some of the participants who had gone through an appeal felt that there was a lack of support from the faculty when an evaluation was challenged. Having support may make them more likely to fail a student or resident.

At the university, when a student fails and s/he writes a letter defending their actions, the letter is sent back to the preceptor . . . please defend your actions in response to this letter. And it's like being accused of not being truthful. . . . We need support for when that happens. That may seem really silly, like, oh I failed a student, boo hoo, I need moral support. But you know? You do.

### Discussion

The findings of this study would seem to suggest that efforts to improve the psychometric properties of the evaluation forms and efforts to teach faculty how to fill out the forms have had only limited success because they have targeted only one area of weakness in the evaluation system. It would appear that faculty do not lack the ability to use the forms as they currently exist, rather they lack the willingness. If ratings are to improve, we must focus not only on the psychometric aspects of the evaluation system, but also on the psychosocial aspects.

This is not to say that supervisors felt no responsibility for reporting problematic students. In fact, participants repeatedly asserted their sense of responsibility to society, the program, and the trainee. However, despite this very strong motivation, several barriers to failing a trainee were noted. Four broad areas were identified: (1) lack of documentation, (2) lack of knowledge of what to specifically document, (3) anticipating an appeal process and (4) lack of remediation options. Given the nature of these barriers, several interventions suggest themselves.

First, efforts must be made to develop mechanisms that enable timely and appropriate documentation of performance. Our participants suggested that technology may be able to facilitate the documentation of trainee performance and reduce the amount of time this takes. However, supervisors still need to know what information to record

when using such technology. A computer program by itself will not resolve all of the documentation issues. Most of the study participants are aware that stating that the resident is lazy and irresponsible on the ITER is inadequate evidence to support a failing rotation but they do not know what should be recorded instead. Carefully directed faculty development efforts toward these issues could resolve these problems much more effectively than changes to the forms.

Second, options must be developed to reduce the cost to the supervisor. As just one example, a resource office and support system could be provided for clinical supervisors to serve many roles. First, it would act as a point of contact for supervisors when they first recognize that they are dealing with a trainee who is failing to meet expectations. The office should be able to clarify what information needs to be collected and what steps need to be taken to comply with the university's evaluation process. This would help to address some of the problems with documentation. Second, this office should be able to counsel the supervisor on how to provide "bad news" to students, present information in a written evaluation to support their impressions, interact with the trainee who challenges their opinion, and handle the appeal process if it occurs. Third, the office should be there to provide support to the supervisor who is going through an appeal process. Situational factors have been found to be more important than personal factors in resisting pressures to conform.<sup>12</sup> The addition of this resource office may change the situation to one that is more supportive of failing trainees who do not meet expectations, and as a result, we may see an increase in the willingness to report poor clinical performance.

Finally, an important barrier to reporting poor performance was the supervisors' concern about not having a good system in place for remediating failing trainees. The supervisors tended to assume the responsibility for creating a remediation plan as opposed to letting the program director take on this role. The impact of a single failing evaluation on the trainee's overall program evaluation also appeared to be overestimated by the study participants. As well, they tended to assume that no remediation was available which may not always have been correct.

This sense of obligation to the trainee may in part be reflective of the supervisors' lack of knowledge regarding whose responsibility it is to provide a remediation program, another area to be clarified in faculty development sessions.

This study is only an initial foray into the social and psychological issues associated with the evaluation of poorly performing trainees. It was conducted at a single institution using volunteer subjects from only two specialties which limits its generalizability. Nonetheless, it is an important first step in understanding the social, psychological and systemic pressures that stop supervisors from expressing the true level of ability that they have determined in their clinical trainees. This information is necessary if we are to develop an evaluation system that enables clinical supervisors to

consistently report poor clinical performance.

This research was supported through grants to the authors from the Royal College of Physicians and Surgeons of Canada and the University of Ottawa. Dr. Regehr is supported as the Richard and Elizabeth Currie Chair in Health Professions Education Research.

## References

- 1 Kassebaum DG, Eaglen RH. Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Acad Med.* 1999;74:842–49.
- 2 Gray JD. Global rating scales in residency education. *Acad Med.* 1996;71(1 suppl):S55–S61.
- 3 Mahara MS. A perspective on clinical evaluation in nursing education. *J Adv Nurs.* 1998;28:1339–46.
- 4 Cohen GS, Blumberg P, Ryan NC, Sullivan PL. Do final grades reflect written qualitative evaluations of student performance? *Teach Learn Med.* 1993;5:10–15.
- 5 Speer AJ, Solomon DJ, Ainsworth MA. An innovative evaluation method in an internal medicine clerkship. *Acad Med.* 1996;71(1 suppl):S76–S78.
- 6 Hatala R, Norman GR. In-training evaluation during an internal medicine clerkship. *Acad Med.* 1999;74(10 suppl):S118–S20.
- 7 Cohen GS, Henry NL, Dodd PE. A self-study of clinical evaluation in the McMaster clerkship. *Med Teach.* 1990;12:265–72.
- 8 Tonesk X, Buchanan RG. An AAMC pilot study by 10 medical schools of clinical evaluation of students. *J Med Educ.* 1987;62:707–18.
- 9 Albanese M. Rating educational quality: factors in the erosion of professional standards. *Acad Med.* 1999;74:652–58.
- 10 Strauss A, Corbin J. *Basics of Qualitative Research.* Thousand Oaks: Sage Publications, 1998.
- 11 Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. *BMJ.* 2000;320:50–52.
- 12 Myers DG, Spencer S. *Social Psychology.* 2nd ed. Toronto: McGraw-Hill Ryerson, 2004.