Motivational Consequences that Affect the Timeliness and Candidness of Resident Evaluations

Michael FitzGerald* and Daniel Schumacher

Cincinnati Children's Hospital Medical Center, Division of Emergency Medicine,Cincinnati, Ohio, USA

Abstract: Motivational factors that undermine the *willingness* of raters to be timely and candid in their evaluation of residents has not been well studied. According to expectancy theory, this motivation stems from perceptions regarding the *likelihood* and *desirability* of various consequences associated with those actions.The aim of this initial investigation is to identify the range of positive andnegative consequences that might drive a rater's motivation toward - or away from - being timely and candid when completing resident evaluations.

Methods: This was an exploratory descriptive investigation using semi-structured interviews with nine physician faculty at Cincinnati Children's Hospital Medical Center. Faculty were asked to describe the potential consequences of being timely (or not) and being candid (or not) in their resident evaluations. The consequences described by respondents were analyzed using a thematic coding process.

Results: A range of consequences were identified including personal-related (e.g., a sense of fulfilling one's role obligations); resident-related (e.g., affecting self-confidence); and institution-related (e.g., chances of renewed accreditation).

Conclusions: The range of potential consequences identified in this investigation shines a light on the motivational factors that should be considered with regards to rater timeliness and candidness and providesthe necessary information for the next step: the development of a tool for assessing raters' perceptions of the *like lihood* and *desirability* of each consequence in a specific rating context.

INTRODUCTION

A number of psychometric factors including poor instrument design and non-conscious error can undermine a rater's *capability* to assess and evaluate medical learners in a timely and accurate way. These factors alone pose a significant challenge and are the targeted focus of most current efforts to improve medical learner evaluations. [1, 2] But raters are not simply passive, error-prone agents in the evaluation process [3]; they make active and conscious decisions regarding timeliness and candor. Any efforts and resources devoted to ensuring the psychometric soundness of an evaluative process are wasted if raters in a given setting are *unwilling* (i.e., unmotivated) to be timely and candid.

The mass shooting at Fort Hood in 2009 by Nidal Halik Hasan provides an extreme but illuminating example of the potential cost of not understanding and addressing motivational factors in the evaluation process. In this case, a number of supervisors held private concerns about Hasan's performance but those same supervisors not only failed to document those concerns on his evaluation but also gave him the highest performance ratings possible [4]. The failure to report negative performance was a conscious decision with

disastrous consequences. Far more pervasive, but no less devastating, is the accidental harm to patients that can result when supervisors are able, but not motivated, to provide timely and candid evaluations.

A significant effort must be made to understand, identify, and address the motivational issues that might undermine a rater's *willingness* to provide timely and accurate performance information in their setting.

Researchers in organizational psychology have conducted a number of investigations on the role of motivational factors in rater leniency (i.e., rater inflation) which is a pervasive problem in performance evaluation. Their research indicates that raters are often motivated to inflate ratings to avoid potential negative consequences including defensive reactions, damaged work relationships, and time-consuming follow-up discussions and requirements.[5, 6] Relatedly, the risk of rater inflation increases when there is a real, or perceived, lack of rater anonymity [7].

While the issue of rater motivation has been acknowledged as a problem by medical educators, [8, 9] only one empirical study has been conducted to identify the specific motivational factors that should be addressed. In this study by Dudek, Marks, & Regehr [10], faculty supervisors were interviewed to "explore their perspectives on the evaluation of poor performing students or residents" and asked to "describe why they might be reluctant to fail a poorly perform-

*Address correspondence to this author at the Cincinnati Children's Hospital Medical Center, Division of Emergency Medicine, Cincinnati, Ohio, USA; Tel: 513-636-1373; Fax: 513-636-7967;
E-mail: mike.fitzgerald@cchmc.org

ing student."Similar to the studies cited from organizational psychology, faculty supervisors said they might be reluctant to fail a poorly performing student because it would require their participation in difficult appeals process. In this case, the appeals process was seen as particularly agonizing because there was rarely any additional documentation available to back up their private judgments.

The existing research provides a good starting point for understanding and addressing rater motivation in medical education but it only points to a narrow range of motivational factors that could undermine rater candidness and doesn't address the issue of rater timeliness.

Murphy and Cleveland proposed a potentially promising approach for (1) identifying the broader range of motivational factors in rater candidness and timeliness and (2) predicting the influence, or motivational power, of each in a specific context. [11] This approach is based on Vroom's expectancy theory of motivation [12] which suggests that raters – like most people in most situations – are motivated to act based on their own perceptions regarding the *likelihood* and *desirability* of various potential consequences. In the case of rater inflation, a given rater is likely to inflate ratings if the perceived consequences of doing so are more desirable than the consequences of giving candid low ratings. Thus, the greater the perceived likelihood and desirability/undesirability of the consequences the greater the motivation to inflate ratings.

The first step in applying this approach in any practical effort to improve rater timeliness and candidness is to identify the range of potential consequences for the rater. The second step is to develop a tool for assessing individual rater's perceptions regarding the likelihood and desirability of eachconsequence. This would provide an indication of the motivational climate (i.e., the degree to which raters in a given setting are motivated to provide timely and candid evaluations) and would point to the specific consequences that should be targeted to maximize rater motivation.

The aim of this initial investigation is to address the first step in the process by identifying rater perceptions regarding the various potential consequences associated with timeliness and candidness in the context of monthly resident evaluations. We think this will provide the necessary information for the next step: the development of a tool for assessing raters' perceptions of the *likelihood* and *desirability* of each of these consequences in a specific rating context.

METHODS

Participants

The population for this investigation was limited to physician faculty at Cincinnati Children's Hospital Medical Center (CCHMC) who are required to provide end-of-rotation evaluations of residents. It was necessary to limit the population of interest to a specific institution and type of evaluation because the consequences of rating behaviors are likely to vary among institutions and evaluations conducted.

A full alphabetically ordered list of eligible faculty names was obtained and reordered based on randomly assigned numbers. A request for participation was sent *via* email to the first twenty faculty members on the list. The request for voluntary participation included a study information sheet describing the purpose of the study, the nature of their participation, and measures taken to protect confidentiality.

Consistent with techniques of theoretical sampling, [13] noprior estimate of sample size was calculated. Rather, data collection continued until it was determined that saturation of the recurrent themes had been achieved. In this case, saturation was reached after conducting nine interviews obtained from a total of eighty five requests sent over a three month period and was consistent with general estimations for the number of interviews needed to achieve saturation. [14] The nine faculty members included two from Emergency Medicine, two from Adolescent Medicine, three from General Pediatrics, one from Cardiology, and one from Critical Care.

Design

This was an exploratory descriptive investigation using semi-structured interviews. The data from this investigation will be used to develop a measure to assess the motivational climate with regards to timeliness and candidness.

Interview Procedure

The interviews were conducted at the location most convenient for the interviewee and lasted no more than thirty minutes. The principal investigator conducted the semi-structured interviews during which each faculty interviewee was asked to consider the issues of timeliness and candidness in resident evaluations. With regard to *timeliness*, the interviewee was asked to assume that they regularly completed their evaluations within two days of the initial request. With regard to *candidness,* the interviewee was asked to assume that they were regularly *candid* in their ratings and didn't deliberately inflate the ratings of poorer performers.

The interviewee was asked to generate their own ideas regarding various *potential* positive and negative consequences that might be experienced by themselves, the resident, the institution and patients: first for being regularly timely and then for being regularly candid. Brief notes were taken during the interview to capture key responses and facilitate the discussion.

Analysis

After each interview, the notes were transcribed into an electronic format. The primary investigator assigned thematic codes to key responses using NVivo 9; [15]qualitative data analysis software that facilitates the coding process. This process continued until– as Strauss has suggested - thematic codes reached a level of "saturation" or a point at which all incoming data readily fit into existing codes and the emergence of new themes ceased.[13]

RESULTS

The potential positive and negative consequences for being regularly timely and candid as reported by interviewees are shown in Table **1**. It should be noted that respondents were also asked to identify negative and positive consequences for *not* being timely or candid. However, to simplify the list, those negative consequences are restated in a positive direction as consequences that could be avoided and the positive consequences are restated as negatives. For example, a potential negative consequence of *not* being timely or *not* being candid was a *decreased* chance of getting a raise. This is reported on the table as an increased chance of getting a raise if one *is* timely or candid. Similarly, the potential positive consequence of avoiding interpersonal conflict when one chooses to inflate ratings is reported as a potential negative when one chooses to be candid.

A number of other consequences mentioned by respondents are not included on this list because they were not potential *direct* consequences of being timely or candid. For example, a number of respondents mentioned that patients might receive better care if residents receive timely and candid information about their performance. This might be true but the consequence of better care is actually linked to the more direct consequence of "an increased likelihood of improving important skills."

While the aim of the interviews was to capture the full range of potential consequences for being timely and candid, many participants also wanted the opportunity to discuss their own views about the likelihood and desirability of the consequences they identified. Almost every respondent mentioned that they knew very little about how the evaluation data was used, or with whom it was shared. As a result, they were dubious about the likelihood of any positive consequences that might result from their evaluations. As one individual stated, "I complete the form and hit send but – as far as I can tell – it just disappears into the ether." Many also expressed some degree of dismay regarding the complete lack of negative consequences for not completing evaluations in a timely manner (or lack of positive consequences for being timely). The chance of receiving a reminder emails was the only real (i.e., likely) negative consequence respondents could identify.

DISCUSSION

Findings

This investigation revealed a range of positive and negative consequences that might affect rater timeliness and candidness when completing monthly resident evaluations and is a critical first step in the development of a practical approach for addressing the neglected issue of rater motivation. Some of the identified personal-related consequences, such as those pertaining to pay and career opportunities, were not surprising, but the interviews did reveal a few consequences that were less expected including "a sense of fulfilling career obligations" and "maintaining a positive image as an educator." The semi-structured design of the interviews effectively ensured non-personal consequences were identified including those that might affect residents (e.g., "decreased self-

Table 1.   Potential Consequences for being Regularly Timely or Regularly Candid in Resident Evaluations

CONSEQUENCES*	FOR TIMELINESS	FOR CANDIDNESS
FOR SELF	• (Pos) Increased sense of fulfilling professional role obligations	
	• (Pos) Increased reputation for being an effective educator	
	• (Pos) Increased chances of getting a raise or other financial reward	
	• (Pos) Increased opportunities for career advancement	
	• (Pos) Avoid reminder notices	
		• (Neg) Interpersonal conflict with the resident
		• (Neg) Receiving a low faculty evaluation rating from the resident
		• (Neg) Having to participate in a remediation process
FOR RESIDENT	(Pos) Increased likelihood of improving important skills	
		• (Neg) Decreased self-confidence
FOR INSTITUTION	(Pos) Increased chances of renewed accreditation	
		• (Pos) Reputation for providing honest feedback.
		• (Neg) Reputation for being too tough on evaluations.
		• (Neg) Reputation for having residents who perform poorly relative to other institutions.

* Note: Shaded consequences apply to both timeliness and candidness.

confidence") and the institution/program (e.g., "increased chances of renewed certification").

Implications

Identifying and understanding the full range of potential consequences alone is valuable; it broadens the lens on the kinds of motivational factors that should be considered when there are problems with rater timeliness and candidness in a given setting. To be of more practical use, however, medical educators will need a diagnostic tool for assessing raters' perceptions of the *likelihood* and *desirability* of each consequence in a specific rating context. The development of this tool is the next step.

Without delving too deeply into the complexities, it is important to convey how assessing the *likelihood* and *desirability* of the consequences identified in this investigation might lead to practical interventions to improve rater motivation. In general terms, to improve rater motivation in a specific context, medical educators should develop targeted interventions to ensure that raters' *most desirable positive consequences* are seen by them as *very likely* to occur and the *least desirable negative consequences* are seen as *very unlikely.*

The types of interventions that might be needed would range from very simple to very complex. As an example of a simple intervention, consider a setting where rater timeliness is a chronic problem that was jeopardizing reaccreditation status. Suppose that one was able to determine that raters truly saw reaccreditation as a highly desirable positive consequence yet were unaware of its link to reaccreditation decisions. In this case, some form of simple communication to clarify the link between evaluation timeliness and reaccreditation could have a strong impact on improving the motivation to be timely.

On the more complex side, suppose that rater inflation (i.e., a lack of rater candidness) was also a problem in this setting and raters believed it was very likely that they would have to participate in a lengthy and complex remediation process (i.e., a highly undesirable negative consequence) if they gave candid low ratings. The effective intervention in this case would require some improvement to make the remediation process shorter and simpler.

In some cases, it might be very difficult to even determine the appropriate intervention strategy. For example, suppose raters in a given context have a very strong desire to avoid negative resident reactions and see it as a very likely consequence of giving a candid low rating. In this case, an effective intervention would require not only an understanding of what negative reaction(s) are feared but also an understanding of the social or organizational factors driving those reactions.

Limitations

While this investigation did reveal a broad range of potential consequences, it should be noted that some potentially relevant consequences might have been missed due to respondent selection bias. The potential interviewees were selected using a random process but the poor response rate increased the likelihood of a self-selection bias. Those who agreed to participate might have been biased in the direction of being positively motivated to be timely and candid in the first place. Consequently, they might have missed some of the relevant negative consequences undermining the motivation of other raters in this setting. In addition, the consequences identified in this investigation were from raters at the same institution in the context of a single evaluation context (i.e., monthly resident evaluations). Consequences relevant at other institutions or in other contexts might have been missed.

Despite these limitations, the consequences identified represent a broad range and should serve as an effective basis for the development of a useful diagnostic tool. To better ensure the inclusion of other relevant consequences, open-ended questions could be included in the design of the tool.

CONCLUSION

Raters make conscious decisions about whether or not to be timely and candid in their evaluation of residents. A lack of rater timeliness and candidness means that residents are not receiving the information they need to improve and programs are not receiving the information they need to address performance problems before they cause harm to patients or colleagues. The path from the consequences identified in this investigation to improving rater motivation might not be simple, but is necessary to effectively address problems undermining rater motivation.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1] Gray JD. Global rating scales in residency education. Acad Med 1996; 71: S55.
[2] Swing SR. Assessing the ACGME general competencies: General considerations and assessment methods. Acad Emerg Med 2002; 9: 1278-88.
[3] Murphy KR, Cleveland J. Understanding performance appraisal: Social, organizational, and goal-based perspectives. USA: Sage Publications Inc 1995.
[4] Fort hood shooting: trouble at every turn in Hasan's training. Idaho Press Tribune. 2010 January 10, 2010; Sect News.
[5] Levy PE, Williams JR. The social context of performance appraisal: A review and framework for the future. J Manage 2004; 30: 881-905.
[6] Villanova P, Bernardin HJ, Dahmus SA, Sims RL. Rater leniency and performance appraisal discomfort. Educ Psychol Meas 1993; 53: 789.
[7] Lerner JS, Tetlock PE. Accounting for the effects of accountability. Psychol Bull 1999; 125: 255-75.
[8] Cohen GS, Blumberg P, Ryan NC, Sullivan PL. Do final grades reflect written qualitative evaluations of student performance? Teach Learn Med 1993; 5: 10-5.
[9] Albanese M. Rating educational quality: Factors in the erosion of professional standards. Acad Med 1999; 74: 652-8.

[10]   Dudek NL, Marks MB, Regehr G. Failure to fail: The perspectives of clinical supervisors. Acad Med 2005; 80(10 Suppl): s84-7.

[11]   Murphy KR, Cleveland J. Rater motivation. In: Understanding performance appraisal: Social, organizational, and goal-based perspectives. Thousand Oaks: Sage Publications Inc 1995: p. 241.

[12]   Vroom VH. Work and motivation. New York: Wiley USA 1964.

[13]   Strauss AL, Corbin JM. Basics of qualitative research: Techniques and procedures for developing grounded theory. USA: Sage Publications Inc 1998.

[14]   Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2006; 18: 59-82.

[15]   NVivo qualitative data analysis software; QSR International Pty Ltd. Version 9, 2010.