A problem-oriented approach to resident performance ratings

Reed G. Williams, PhD,[a] **John D. Mellinger, MD,**[b] *and* **Gary L. Dunnington, MD,**[a] *Indianapolis, IN, and Springfield, IL*

Background. *Global, end-of-rotation evaluations are often difficult to interpret due to their high level of abstraction (eg, excellent, good, poor) and the bias toward high ratings. This study documents the utility of and measurement characteristics of serious problem items, an alternative item format.*
Methods. *This report is based on 4,234 faculty performance ratings for 105 general surgery residents. Faculty members reported whether each resident had a serious problem for each of 8 areas of clinical performance and 6 areas of professional behavior.*
Results. *A total of 263 serious problems were reported. The performance category with the most total serious problems noted was knowledge and that with the least problems noted was relations with patients and family members. Seven residents accounted for 86.9% of all serious problem reports. Each resident had serious problems in multiple performance areas. Problems were reported most frequently in knowledge, management, technical/procedural skills, ability to assume responsibility within level of competence, and problem identification. Citations of these serious problems were most common in year 3. Traditional ratings of global performance were not an adequate means for identifying residents with serious performance problems.*
Conclusion. *Serious problem ratings can communicate faculty concerns about residents more directly and can be used as a complement to conventional global rating scales without substantially increasing faculty workload. (Surgery 2016;160:936-45.)*

From the Department of Surgery,[a] *Indiana University School of Medicine, Indianapolis, IN; and the Department of Surgery,*[b] *Southern Illinois University School of Medicine, Springfield, IL*

> *"When the meaning is unclear there is no meaning"*[1]
>
> —*Marty Rubin*

VIRTUALLY ALL RESIDENCY PROGRAMS use an end-of-rotation, global rating form as a primary element of their system of resident assessment. These forms require faculty supervisors most commonly to summate their impressions of overall resident performance throughout the rotation and translate those impressions and judgments into performance ratings (eg, knowledge, clinical performance, professionalism) using standard items and quality rating anchors (eg, excellent, very good, good, fair, poor) describing facets of performance.

Presented at the 73rd Annual Meeting of the Central Surgical Association, Friday, March 11, 2016, Montreal, Quebec, Canada.

Accepted for publication April 13, 2016.

Reprint requests: Reed G. Williams, PhD, Department of Surgery, Indiana University School of Medicine, Indianapolis, IN 46202. E-mail: reedwill@iupui.edu.

0039-6060/$ - see front matter

http://dx.doi.org/10.1016/j.surg.2016.04.040

Use of these forms has proven problematic over many years despite efforts to solve the problems.[2] Initial efforts were focused on altering instrument formats (checklists, rating scales, rating scales with descriptive anchors). Later efforts were focused on training users to make more effective use of the forms. Neither method solved the problems. More recently, the focus has been on better understanding of the rater and altering the content of rating items and quality rating anchors to fit the rater's frame of reference.[3-7]

To understand raters' frames of reference, Ginsburg et al[6] had internal medicine training supervisors recall an outstanding resident, an average resident, and a resident with performance problems and describe each resident. The investigators did not structure or otherwise influence the descriptions provided by participating supervisors. Supervisors' descriptions did not align well with the forms used typically to rate residents. Specifically, supervisors started with a single performance characteristic that defined each resident and built the overall characterization around that performance characteristic.

Crossley et al[3] modified the content of 3 existing scales of performance rating to better align

with the way that supervisors think about medical trainees. By using the construct-aligned scales in parallel with conventional versions to assess approximately 2,000 medical trainees, they found that construct-aligned scales decreased examiner disagreement and increased the ability of the instruments to discriminate among trainees. In their discussion of study limitations, they stated that their behavior descriptors could "probably be improved," suggesting that efforts to make pairs of comparable instruments led to an admixture of characteristics of traditional and new items. New procedures of performance rating that are designed exclusively to align better with the way resident supervisors think about medical trainees may produce even greater effects.

Our base form of clinical performance and professional behavior had 3 items asking for global judgments about clinical performance, professional behavior, and performance relative to that of other residents. In an earlier publication,[8] we provided evidence about the psychometric properties of this scale. We believe that the 3-item scale lacks resolution when it comes to identifying specific performance problems that need attention.

To address this concern, we added a "serious problem feature" to our system of performance rating. The specific changes implemented were inspired by and modeled after a procedure used by Consumer Reports to elicit the self-reports of car owners about serious problems with the cars they own (eg, Consumer Reports, April 2013, p. 86[9]). Consumer Reports asks raters to identify trouble spots in key operating systems of their cars (eg, transmission, cooling system, electrical system) by giving them a list of key operating systems and asking them to identify systems where there have been serious problems. Owners use their own definition of serious problems. As such, the items ask the owners to report operational judgments that they make in everyday life based on their experience with their car.

We felt that ratings like these have more direct meaning than do the quality ratings used normally to characterize resident performance. These ratings stimulate the rater to make a black or white decision about facets of resident performance unlike that required by typical scaled evaluations, which both soften the rater's critique and lack precision.

Our current study describes the adaptation of the Consumer Report procedure that we introduced into one general surgery residency program in 2006 and provides results based on data from academic years 2006–2007 through 2012–2013. The results are tested against the following 4 criteria: (1) The results should identify specific problems in the performance of trainees. (2) The results should differentiate trainees with serious performance problems from those with no performance problems. (3) Examiners should agree with one another in their identification of serious problems of performance of residents. (4) The assessment process should be better aligned with the way attending surgeons think about residents.

METHODS

Clinical performance and professional behavior form. In 2001, we created and began using the Clinical Performance and Professional Behavior form in a single general surgery residency program. This work was informed by the findings of Verhulst et al.[2] The form required expert rater assessments using 3 items. One asked for a holistic impression of clinical performance. The second asked for a holistic impression of professional behavior, and the third asked the rater to judge the resident's performance compared to the performance of other residents at the same level of training, hereafter called comparison ratings. We have described the rationale for and psychometric performance of these items in an earlier publication.[3]

Serious problem items. Since 2006–2007, faculty members also have been asked to indicate whether the resident had what the faculty member considered a serious problem in any of 8 areas of clinical performance (data collection, problem identification, diagnostic approach, management, knowledge, self-directed learning habits, technical/procedural skills, and intraoperative decision-making) or the 6 areas of professional behavior performance (communication, relations with patients and family members, relations with other medical personnel, reliability and dependability, ability to assume responsibility within level of competence [neither under or over confident], and equanimity).

The item reads as follows: "Check any of the following clinical performance (professional behavior) attributes that you consider a serious problem for this resident at this time." Faculty responded by checking a box for those performance characteristics that were considered a serious problem. Faculty members defined serious problems for themselves. The specific performance areas were also patterned after the work of Verhulst et al.[2]

Rating process. All faculty members on a service were invited to rate residents who were on that service