The Feasibility and Acceptability of Implementing Formal Evaluation Sessions and Using Descriptive Vocabulary to Assess Student Performance on a Clinical Clerkship

Michael J. Battistone
Department of Medicine
Veterans Affairs Salt Lake City Health Care Center
and
University of Utah School of Medicine
Salt Lake City, Utah, USA

Caroline Milne and Merle A. Sande
Department of Medicine
University of Utah School of Medicine
Salt Lake City, Utah, USA

Louis N. Pangaro and Paul A. Hemmer
Uniformed Services University
F. Edward Hebert School of Medicine
Bethesda, Maryland, USA

T. Samuel Shomaker
University of Hawaii
John A. Burns School of Medicine
Honolulu, Hawaii, USA

Background: A systematized approach to descriptive evaluation of clinical performance using a vocabulary of global descriptors in the setting of formal evaluation and feedback sessions has been shown to be reliable and valid. The feasibility of this method beyond the institution at which it was developed has not been studied.

Purpose: To determine the feasibility and acceptability of implementing formal evaluation and feedback sessions, using a vocabulary of global descriptors, in a third-year core clinical clerkship.

Methods: In 1997, the University of Utah internal medicine clerkship introduced an evaluation method in which student performance was discussed at formal sessions, using a taxonomy of global terms describing progressive development from “reporter” to “intermediary” to “manager/educator” (R–I–M–E). The sessions were face-to-face meetings between the clinical teachers and a clerkship director, at three-week intervals through the twelve-week clerkship at the inpatient teaching sites. Following the evaluation session students met individually with the clerkship director for feedback. To determine feasibility, the authors estimated the time and resources necessary to administer the system and recorded actual teacher attendance over 2 academic years (1997–99). Anonymous surveys, using a four-point, Likert-type scale, were used to determine acceptance of the method by faculty, residents, and students.
**Results:** Attendance was high for residents (79%) and faculty (72%). Mean survey responses from residents and faculty rated the descriptive system “more valid” than the previous method. A majority of the students rated the method as either “helpful” (30%) or “very helpful” (50%). Time requirement for eight to ten students at each teaching site, for evaluation and feedback sessions was one-half day per week of the clerkship director, every three weeks.

**Conclusion:** Our experience establishes the feasibility of implementing this system of formal evaluation and feedback, using descriptive vocabulary, beyond the institution at which it was developed. Students, residents, and faculty endorsed this evaluation system and the survey results suggest substantial utility and face validity.

Education, in order to be effective, must include valid and reliable methods of assessment. These methods must also be feasible, and ideally, acceptable to those responsible for using them and to those whose work is being evaluated. Regardless of other curricular issues, the credibility of both formative and summative evaluations is indispensable. The quality of evaluations that students receive shapes, if not defines, the quality of their clinical training.

There have been significant differences in techniques of assessing performance in clinical clerkships, and concerns have been expressed over the relative lack of research in this field. Approaches have included: numeric rating scales, sets of standardized comments, and descriptors of student abilities on evaluation forms. Most clerkship evaluation forms use an “analytic” model in which values, attitudes, skills, and knowledge are considered individually, while others seek a “global” evaluation, in which many domains of performance are subsumed by a single rating. Regardless of the differences, threats to the validity of evaluations are common at all institutions. Clerkship directors confront problems of insufficient observation, lack of meaningful comments by teachers, inconsistency in evaluation across sites, and inter- and intra-observer variability.

To address these threats, a method of formal evaluation and feedback sessions was introduced at Uniformed Services University of the Health Sciences F. Edward Hebert School of Medicine (USUHS) two decades ago. Subsequently, a systematic set of global terms describing progressive student performance (Reporter–Interpreter–Manager/Educator, or R–I–M–E) was developed by Pangaro and has been used effectively in the internal medicine clerkship. The reliability of this system has been shown to be greater than 0.8, comparable to quantifiable examinations, and sufficient for high stakes decisions. Predictive validity has been demonstrated for identifying students at risk of failing an end-of-clerkship National Board of Medical Examiners medicine subject exam and for forecasting low scores for intern performance. Originally reported to be valuable in identifying students with academic difficulties, this approach is also useful in assessing professionalism. Finally, this evaluation method also provides “real-time” faculty development for housestaff and faculty. Despite this evidence, studies of feasibility of this method across other institutions or disciplines appear not to have been published.

In 1997, the University of Utah School of Medicine began implementing a revised curriculum. While some schools consider reforms of evaluation methods after other aspects of the curriculum have been changed, we chose to make this a top priority. In an effort to improve both formative and summative assessment of students in the internal medicine clerkship, formal evaluation and feedback sessions, using the vocabulary developed by Pangaro, were introduced. This paper describes our experience in adopting these methods.

**Methods**

**R–I–M–E Vocabulary**

Performance goals for students were described by the following terms:

- **Observer:** A student in pre-reporter status, not meaningfully contributing to patient care activities (we at the University of Utah added this descriptor to the original vocabulary).
- **Reporter:** Denotes reliability and competence in collecting and communicating clinical information, specifically emphasizing the reliability, honesty, and professional qualities required in medical interviewing, physical examination, oral presentations, note writing, follow-through on assigned tasks, and working with patients and hospital personnel.
- **Interpreter:** Describes additional attributes which build on those seen in the “reporter.” This term involves functions of independent, critical thinking in considering clinical data and advocating or refuting diagnostic hypotheses. Students at this stage demonstrate consistency in prioritizing problem lists and offering reasonable differential diagnoses without prodding.
Manager: Describes a high degree of direct involvement in patient care—the level expected of interns and sub-interns. Students who are consistent “managers” can propose diagnostic and therapeutic options, demonstrate judgement in working with patients and families, and reliability in implementing the plan.

Educator: Denotes a pattern of self-directed learning beyond the basics, and frequent contribution to the education of fellow students, residents, and even faculty. Students who function as “educators” demonstrate ability to frame important questions posed by complicated cases, and to resolve these questions through independent and efficient literature research.

Evaluation Sessions

Formal sessions were scheduled regularly through the 12-week internal medicine clerkship at each of three inpatient sites: the University of Utah Hospital, an affiliated Veterans Affairs Medical Center, and a university-affiliated, privately operated community hospital. Evaluation sessions were not scheduled to review performance in the ambulatory portion of the course. The conferences were scheduled to coincide with the conclusion of the faculty tour of duty on ward service—this was every four weeks during the first year of implementation, but was revised to every three weeks, reflecting a change in the duration of faculty assignments. Sessions were conducted in the same conference room used daily for morning report, a private area far from the patient wards. The evaluation sessions were moderated by the clerkship director (or assistant clerkship director) and attended by ward residents (postgraduate year-2 [PGY–2] and PGY–3) and faculty on service. Each evaluator was asked to verbally describe their students’ performance using the R–I–M–E vocabulary, and to provide specific observations to substantiate their rating. Evaluators were directed to conclude their critiques by identifying a specific “next step” to which students’ future efforts might be directed, and to turn in their written evaluations before leaving the session. Faculty had been briefly introduced to the definitions and use of the R–I–M–E vocabulary at two department meetings and these points were briefly reviewed at each evaluation conference. In addition to the teaching faculty and residents, the medicine chair (or designee) and chief medical resident attended the session, and added their own observation of students’ performance in professor rounds and in the weekly chief resident EKG conference. These comments were offered only after evaluations from the ward team had been presented, and were not framed in terms of the R–I–M–E vocabulary. Each session was scheduled for one hour, and reviewed the performance of eight to ten students.

Feedback Sessions

In the three hours following the evaluation session, eight to ten students met individually with the clerkship director to review the written evaluations, as well as the spoken comments offered by the resident, attending, chief resident, and chair. Each student appointment was anticipated to last about fifteen minutes. Emphasis was given to how the student could achieve the prescribed “next step” consistent with the department’s goals for students.

Outcome Measures

To determine whether this approach to student assessment would be feasible at our institution, we recorded, for two consecutive clerkship years (1997–99), resident and faculty attendance at the evaluation sessions. We also estimated the time required of the clerkship director and administrator to conduct the evaluation and feedback sessions.

Table 1. Attending and Resident Survey and Descriptive Statistics

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>Res</th>
<th>Att</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>How useful are the formal evaluation sessions in evaluating students?</td>
<td>23</td>
<td>71</td>
<td>3.3</td>
<td>.6</td>
</tr>
<tr>
<td>How useful is the R–I–M–E vocabulary in evaluating students?</td>
<td>23</td>
<td>70</td>
<td>3.4</td>
<td>.7</td>
</tr>
<tr>
<td>How useful is the R–I–M–E vocabulary in giving feedback to students?</td>
<td>23</td>
<td>70</td>
<td>3.2</td>
<td>.7</td>
</tr>
<tr>
<td>How would you rate the validity of our current method of student evaluation compared with our previous system or other systems with which you have experience?</td>
<td>23</td>
<td>67</td>
<td>3.3</td>
<td>.7</td>
</tr>
</tbody>
</table>

Note:  Res = resident; Att = attending. Results from an anonymous survey of 85 faculty and 67 PGY–2 and PGY–3 residents at the University of Utah, during academic year 1998–1999. Response rate was 84% for faculty and 34% for residents. Each item was scored on a 4-point, Likert-type scale, anchored at the extremes of range, from 1 (not very useful) to 4 (very useful) for items 1–3 and 1 (much less valid) to 4 (much more valid) for item 4. The survey was developed to seek instructor opinion regarding a newly implemented evaluation system on the internal medicine clerkship, which consisted of formal evaluation sessions, feedback sessions, and a descriptive vocabulary of student performance known as R–I–M–E (Reporter–Interpreter–Manager–Educator). The results demonstrate substantial endorsement of the new evaluation system by the faculty and housestaff.
To supplement attendance as a measure of acceptability of this evaluation method, residents and faculty were surveyed anonymously at the end of the second year of implementation (1998–99). The survey instrument consisted of four items, with a four-point Likert-type scale to record each response (Table 1). Surveys were mailed to 67 residents and 85 faculty. Although the surveys were anonymous, resident and faculty respondents were asked to mail (in a separate envelope) a signed statement indicating that they had completed the survey.

In 1998–99, we collected anonymous surveys from students (Table 2) on the final day of their rotation, after they had reviewed their evaluations, but before final grades had been prepared or reported.

**Results**

Forty-eight evaluation sessions were scheduled for the first year and thirty-six for the second. 100% were held as planned. Fewer sessions were scheduled in the second year, reflecting a change in the structure of the ward and ambulatory assignments in the clerkship. Attendance rates in the first year for residents and faculty were 87% and 65%, respectively. In the second year, resident attendance remained high at 82%, and the faculty rate rose to 78%.

Conducting each evaluation session required sixty minutes (average 6–7.5 minutes per student) from the clerkship director every three weeks. Student feedback, given individually at private meetings beginning after the evaluation session had concluded, typically required about 20 minutes per student, or about three hours each three weeks for the director to meet with each student who had been evaluated. An administrative assistant at each site helped distribute evaluation forms, notified residents and faculty of the evaluation session times, and transcribed written evaluations to the final evaluation narrative that was eventually submitted to the dean’s office.

Figure 1 shows the frequency distribution of responses to an item from survey of residents and faculty regarding the implementation of formal evaluation sessions and R–I–M–E vocabulary. Results from the other three survey questions, as described in Table 1, followed similar distribution as that depicted above.

Following a single mailing of the instructor survey, questionnaires were returned by 23 residents (34%) and 71 faculty (84%). Although the surveys were anonymous, it was possible to determine from the signed statements of survey completion that the return rate was greater from PGY–3 residents (41%) than from PGY–2 residents (28%). Figure 2 presents results from the survey item addressing the face validity of the evaluation sessions and R–I–M–E vocabulary. The results demonstrate that responding residents and faculty accepted that the new evaluation system was an improvement over our prior system of evaluation. For the other three survey items, the distribution of responses was similar to that in Figure 2, for both residents and faculty. Six faculty members (out of 71 who returned the survey) rated the system as “much less valid.” From a class of 99 students, 96 (97%) anonymous sur-
veys were completed and returned. Descriptive statistics for each survey item are shown in Table 1. Overall, the students’ responses consistently endorsed the new evaluation system. Of students completing the survey, 51 (53%) reported that their residents used the R–I–M–E terminology in providing personal feedback; 45 students (47%) indicated that their attendings had used the vocabulary in giving feedback.

**Discussion**

Although instructors’ descriptive evaluations of medical students on clinical clerkships are all too often dismissed as “subjective,” there exists a rigorous, reliable, and valid method of evaluating a medical student’s clinical performance that uses synthetic, global descriptors (R–I–M–E) and formal, face-to-face meetings with clinical instructors. Ours is the first study to demonstrate that this approach to evaluating medical students is also feasible and accepted at a medical school beyond that in which the system was developed, and that it can be adopted with substantial face validity and strong support from students, residents, and faculty.

Using this system of descriptive vocabulary with formal evaluation and feedback sessions offers many advantages. First, this approach provides structure across sites and across evaluators, and drives assessment from course expectations rather than preferences of individual raters. Second, this provides a forum for continuous orientation, in which clerkship goals and objectives are frequently articulated and reviewed with both students and teachers. We believe that this promotes a climate that fosters individual student progress and serves missions of resident and faculty development. Indeed, the regular attendance at formal evaluation sessions has in itself been proposed as a case-based faculty development process. Third, this facilitates an essential administrative function for the clerkship director, as complete evaluations are collected without delay, and timely feedback to students is ensured. Finally, and perhaps most importantly, these methods are a powerful method to increase communication among students, residents, faculty, and clerkship directors.

Faculty and housestaff attendance rates establish the feasibility of the method. The response rate for faculty surveys was secondary to the actual attendance rates as an outcome measure; in view of the mean ratings this provides further evidence of the attendings’ support of descriptive evaluations. Unfortunately, the response rate for resident surveys was substantially lower. Although the resident responses were similar to those provided by faculty, our conclusions regarding resident attitudes are limited by the possibility of non-response bias, with failure to collect data from those who felt that the system of descriptive evaluation was much less valid than other approaches. Because this was an anonymous survey, it is not possible to investigate relationships between non-responders and attendance at evaluation sessions, or other aspects of the patterns of their evaluations. Nevertheless, their responses are encouraging.

Students were also pleased with the evaluation system. Most felt that the R–I–M–E vocabulary helped them understand performance expectations and that the feedback sessions during the clerkship were particularly helpful. They felt that such formal feedback sessions would be valuable if used on other clerkships as well. Students rated the one-on-one feedback given by attendings and residents less highly, but students may not have recognized teachers’ comments as feedback since it may not have been labeled as such or included specific ways in which to improve. Attendings and residents were asked to use the R–I–M–E vocabulary in providing feedback to students throughout the course, but there was no formal mechanism to reinforce these instructions prior to the evaluation sessions. These issues were beyond the scope of this study. Importantly, though, even a short time after implementing the evaluation system, nearly half of all students recognized that their residents and attendings were using the R–I–M–E language as part of their feedback to the students, providing further substantiation of the acceptance of the system by instructors.

This system requires a significant investment of time and effort from the clerkship director and his/her administrative staff. However, the time spent in the evaluation process during the clerkship is similar to the time many internal medicine clerkship directors spend collecting, compiling, and assessing evaluations after the clerkship. Although not quantified as part of this study, we did note that calculation of final grades for students was facilitated by the prior collection of evaluation data. Furthermore, the time and resources required for administering this evaluation system are consistent with the expectations of and for the medicine clerkship director, as endorsed by the Association.
of Professors of Medicine.12 Although not part of the system as developed at USUHS, we believe the attendance of the department chair or representative at the sessions served to communicate the importance of these meetings and task of evaluating clinical performance. Accumulated evidence justifies this “real-time” time investment in credible evaluation, as well as the additional benefits of guaranteed feedback to students and sustained development for residents and faculty in their teaching roles.2,3,5,10

There are limitations to this study. We only examined the inpatient portion of the internal medicine clerkship, and continuing studies of this evaluation system in the ambulatory setting are needed. Responses to survey questions are subject to bias. We addressed this issue by making all of the surveys anonymous, but the low resident response rate raises the possibility of selection bias. Nevertheless, it would appear to be unexpected for the non-responding residents to have a dramatically different view of the evaluation and feedback sessions and the R–I–M–E vocabulary given the strong endorsement by the faculty and students. Finally, this study was one of feasibility and acceptability of the evaluation system and did not address specific student outcomes, such as changes in clerkship grade distribution or the detail of teachers’ comments. Outcomes data do exist for this evaluation system,7–9 and the most appropriate first step for adopting this evaluation system was to examine its impact on students and teachers. Currently, studies are underway to address specific student outcomes.

In conclusion, our experience demonstrates that a rigorous and credible system of descriptive evaluation of medical students on clinical clerkships, which has been shown to be reliable and valid, is also feasible at institutions beyond which it was developed. It can be implemented and replace existing methods of student assessment with strong support from students, residents, and faculty. Further research needs to determine whether reliability and validity can be preserved across multiple clerkship settings, or across clerkship disciplines.

References


Received 26 January 2001
Final revision received 2 July 2001