



Group Assessments of Resident Physicians Improve Reliability and Decrease Halo Error

Matthew R. Thomas, MD¹, Thomas J. Beckman, MD², Karen F. Mauck, MD², Stephen S. Cha, MS³, and Kris G. Thomas, MD¹

¹Division of Primary Care Internal Medicine, Mayo Clinic, Rochester, MN, USA; ²Division of General Internal Medicine, Section of Medical Education, Mayo Clinic, Rochester, MN, USA; ³Department of Biomedical Statistics and Informatics, Rochester, MN, USA.

BACKGROUND: Individual faculty assessments of resident competency are complicated by inconsistent application of standards, lack of reliability, and the "halo" effect.

OBJECTIVE: We determined whether the addition of faculty group assessments of residents in an ambulatory clinic, compared with individual faculty-of-resident assessments alone, have better reliability and reduced halo effects.

DESIGN: This prospective, longitudinal study was performed in the outpatient continuity clinics of a large internal medicine residency program.

MAIN MEASURES: Faculty-on-resident and group faculty-on-resident assessment scores were used for comparison.

KEY RESULTS: Overall mean scores were significantly higher for group than individual assessments (3.92 ± 0.51 vs. 3.83 ± 0.38 , $p=0.0001$). Overall inter-rater reliability increased when combining group and individual assessments compared to individual assessments alone (intraclass correlation coefficient, 95% CI=0.828, 0.785–0.866 vs. 0.749, 0.686–0.804). Inter-item correlations were less for group (0.49) than individual (0.68) assessments.

CONCLUSIONS: This study demonstrates improved inter-rater reliability and reduced range restriction (halo effect) of resident assessment across multiple performance domains by adding the group assessment method to traditional individual faculty-on-resident assessment. This feasible model could help graduate medical education programs achieve more reliable and discriminating resident assessments.

KEY WORDS: graduate medical education; professional competence; educational measurement; formal evaluation sessions.

J Gen Intern Med (26)7:759–64

DOI: 10.1007/s11606-011-1670-4

© Society of General Internal Medicine 2011

INTRODUCTION

Residency programs must assess resident competency in patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice.¹ Although reliable and valid assessments of medical knowledge (e.g., United States Medical Licensing Examination, American Board of Internal Medicine certification examination) are readily available, assessment of the other competencies is more difficult.

Faculty ratings are commonly used to assess residents;^{2,3} however, ratings of resident competence by individual faculty are often problematic.⁴ For example, unique post-rotation performance ratings often neglect a large percentage of deficiencies apparent at summative, end-of-year progress judgments.⁵ Ratings forms are sometimes inaccurate and provide little specific formative feedback to residents.⁶ It is difficult to assure the consistent application of performance standards across a large number of faculty members. The problem of leniency has been clearly documented.^{7,8} Lastly, we recognize the so-called "halo" effect, initially described by Thorndike,⁹ in which a good or bad performance in one area affects assessments in other performance domains.¹⁰

Compared with individual faculty assessments, group discussions of clinical performance have proven beneficial in undergraduate medical education. Group discussions have been shown to identify students with marginal funds of knowledge^{11,12} and deficiencies in professionalism,¹³ while improving the quality of narrative comments and better justifying assigned grades.¹⁴ Furthermore, group discussions of learner performance provide a forum for case-based faculty development.¹³ Despite the known advantages of group assessment in undergraduate medical education, little is known about the use of group assessment in residency education.

The aims of this study were to determine (1) whether the addition of faculty group assessments of residents in an ambulatory clinic, compared with individual faculty-of-resident assessments alone, have better reliability and a reduced halo effect; and (2) faculty perceptions of the group assessment process.

METHODS

Setting and Design

This was a prospective longitudinal study performed over the course of one academic year at a large internal medicine residency program. The Mayo Clinic Rochester Internal Medicine Resident Continuity Clinic is organized into 6 firms, each

Received August 6, 2010

Revised January 19, 2011

Accepted February 11, 2011

Published online March 3, 2011

with 24 residents and a mode of 8 (range 7–11) faculty preceptors. During this study each faculty member precepted on one to two specific afternoons each week (a fixed schedule), while the once weekly clinic day of individual residents varied, based on their hospital call schedule (Fig. 1). Therefore, each resident will presented a few cases to several faculty members over the span of several months.

Data Collection

Individual Faculty-on-Resident Assessments. Firm faculty preceptors were selected based on their desire to teach in the continuity clinics and their teaching excellence, which in turn is based on validated teaching assessment ratings.¹⁵ Individual faculty-on-resident assessments were independently completed by firm faculty preceptors on a quarterly basis. Previous validation research on Mayo Clinic internal medicine person-on-person assessments revealed excellent internal consistency reliability and a single dimension of clinical performance based on factor analysis.¹⁶ An average of 4.7 (range 4–8) resident forms (matching those residents scheduled for the upcoming group assessment) were completed quarterly by individual faculty members who were instructed to document direct observations of residents and to decline assessing residents with whom they had not worked. Assessment items, structured on 5-point scales, addressed residents' performances in the following domains (linked ACGME competency in parentheses): accuracy and completeness of data gathering (patient care), effectiveness of interpretation (patient care), appropriate test selection (patient care), illness management (patient care), provision of follow-up (patient care), patient recruitment and retention (communication), and commitment to education (professionalism). Anchors (1 = needs improvement, 3 = average, 5 = top 10%) were provided for each performance domain. Individual assessments were electronically submitted prior to the group assessment.

Group Assessments. Group assessment sessions in each of the six firms were scheduled quarterly (Fig. 1) for 90 min overlapping the usual lunch hour, facilitated by the respective firm chief (a faculty member who administrates for each firm) and attended

by firm faculty (who had submitted individual faculty-on-resident assessments prior to the group assessment). Reviews of six to eight residents, grouped by year of training, were scheduled for each session, which allowed 10–15 min of discussion per resident. Following group discussion of each resident, a single group faculty-on-resident assessment form, with questions identical to those on the individual faculty-on-resident assessment forms, was electronically submitted for each resident by the firm chief, with scores and narrative comments that reflected the consensus of the firm preceptors.

Group assessments were added to the existing components of resident assessment in the continuity clinics, which were left unchanged. Residents are evaluated quarterly in the clinic setting. Continuity clinic advisors (who serve as preceptors in the resident's firm) meet quarterly with residents, during which time the resident's performance in clinic is reviewed.

Faculty Opinion Survey. Faculty opinion surveys were completed immediately following each of the quarterly group assessment sessions. The survey, which was not pre-validated, solicited faculty opinions on whether they agreed (scale 1–5) the group assessment method had improved faculty members' knowledge of each resident's strengths, weaknesses, and learning plan; improved their understanding of assessment in general and confidence in their own assessment skills; and improved resident assessment in the continuity clinic setting. Faculty members' overall satisfaction with the group assessment method was then queried.

Statistical Analysis. Wilcoxon sign rank test was used to compare mean scores, for each item and overall, for assessments by groups and individual raters. Interclass/intraclass correlation coefficient (ICC) was calculated using an ANOVA model where evaluator, evaluatee, and their interaction term were considered. The ICC and its 95% confidence interval were calculated using a random sample from 49 evaluators for individual assessments and 6 evaluators for group assessment. ICC was interpreted as follows: <0.4 poor; 0.4 to 0.75, fair to good; and >0.75 excellent.¹⁷ Halo error, defined in this context as giving similar or identical scores across all item domains, was

Model of a Four-week Calendar for Faculty and Residents in a Single Firm																					
Day	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F	M	T	W	Th	F	
Faculty	A	C	E	G	I	A	C	E	G	I	A	C	E	G	I	A	C	E	G	I	
	B	D	F	H	J	B	D	F	H	J	B	D	F	H	J	B	D	F	H	J	
Resident	1	2	3	4	5	2	3	4	5	1	3	4	5	1	2	4	5	1	2	3	
	6	7	8	9	10	7	8	9	10	6	8	9	10	6	7	9	10	6	7	8	
	11	12	13	14	15	12	13	14	15	11	13	14	15	11	12	14	15	11	12	13	
	16	17	18	19	20	17	18	19	20	16	18	19	20	16	17	19	20	16	17	18	
	21	22	23	24	22	23	24	21	23	24	21	22	24	21	22	23					
Group Assessment Annual Calendar																					
October					January					April					July						
Faculty: All Residents: 1-8 (PGY-1s)					Faculty: All Residents: 17-24 (PGY-3s)					Faculty: All Residents: 9-16 (PGY-2s)					Faculty: All Residents: 1-8 (new PGY-2s)						

Figure 1. Model of a 4-week calendar for a single firm and annual calendar of group assessments. The top part of the figure conceptually illustrates a 4-week calendar for a single firm. Faculty members (indicated by the letters A through J) precept clinic on a fixed day every week. For example, Faculty A always precepts on Mondays. In contrast, residents (indicated by numbers 1 through 24) attend clinic on a different day each week, since their clinic calendar is dependent on their hospital call schedule, which, in turn, is independent of the day of the week. For example, Resident 1 attends clinic on Monday during the first week, and then Friday during the second week. This leads to numerous faculty members gaining experience with a specific resident over time, but no single faculty member who has extensive experience with any specific resident. The lower part of the figure illustrates the annual calendar of group assessments, wherein the performance of residents from a single PGY year is discussed on a quarterly basis in a rotating fashion.

determined by calculating inter-item correlations, which has been used in previous education research¹⁸ and is the most traditional method.^{19,20} Higher inter-item correlations were interpreted to reflect higher halo error, and lower inter-item correlations represented lower halo error. Mean scores were determined for faculty survey responses.

RESULTS

A total of 679 individual faculty-on-resident assessments and 136 group faculty-on resident assessment forms (representing 94% of the group assessment discussions) were collected over a 1-year period. Faculty opinion data were collected from 16 of 18 (89%) sessions.

Scores on items for individual and group assessments are shown in Table 1. Group assessment mean scores on 3 of 7 individual items and the overall average of item mean scores were significantly higher in the group assessments compared with individual assessments (3.92±0.51 vs. 3.83±0.38, p=0.0001). Item inter-rater reliability of individual assessments alone and with the addition of group assessments are shown in Table 2. Overall inter-rater reliability increased when combining group and individual assessments compared to individual assessments alone (intraclass correlation coefficient, 95% CI=0.828, 0.785–0.866 vs. 0.749, 0.686–0.804). Inter-item correlation was less for group (0.49) than individual (0.68) assessments (0.49 vs. 0.68), signifying reduced halo error.

Faculty opinions regarding the group assessments are shown in Figure 2. In general, faculty agreed (on a 5-point scale) that the group assessments improved their knowledge of resident strengths (4.4±0.7), weaknesses (4.5±0.7), and learning plans (3.9±0.9); and improved their understanding of resident assessment (3.8±1.1) and confidence in their assessment skills (3.7±0.9). Faculty strongly agreed that group assessments improved resident assessment in the continuity clinic setting (4.6±0.7) and highly rated their overall satisfaction with the group assessment method (4.3±0.9).

DISCUSSION

Residency programs are charged with accurately assessing resident competency across multiple domains for both formative and summative purposes,¹ yet this task is challenging for

many reasons. Individual faculty members have only brief contact and limited exposure to resident behaviors, so they may miss significant deficiencies in performance.⁵ Additionally, faculty members often provide minimal feedback²¹ and may place variable emphasis on different performance criteria.^{7,10} We report an integrative faculty group assessment model for internal medicine continuity clinics that uses pooled faculty observations to identify learner deficiencies, enable more specific feedback, and provide ongoing case-based faculty development. This study demonstrates improved inter-rater reliability and reduced range restriction (halo effect) of resident assessment across multiple performance domains by incorporating the group assessment method.

Evaluating residents is becoming more challenging. Performance criteria are increasingly explicit, which makes assessments more complex. Faculty practice burdens are growing. Recent recommendations to extend resident duty hour restrictions,²² while potentially important for patient safety, may further reduce resident-faculty continuity. Fortunately, the group assessment model can mitigate these challenges because pooled individual assessment tools² enable faculty to "connect the dots" between separate brief observations in order to discern the larger pattern of performance. Group leaders can ensure that faculty members within a group apply consistent performance standards across many residents. Over time, the group assessment model serves as case-based faculty development with the potential to improve faculty members' knowledge of specific residents, the assessment process in general, and confidence in resident assessment. This type of group assessment model can also be an important component of outcomes-based assessment within residency programs.²³ Properly used, the information gleaned from group assessments may provide residents with more discriminating formative feedback so they can remedy deficiencies and become better physicians.

The use of multiple assessment modalities can potentially minimize deficiencies in any one method.^{2,3,24} Although faculty ratings are the most common method for assessing residents' clinical performance,² this method is fraught with rater errors,⁴ the most common of which is the "halo effect," demonstrated by the tendency of faculty raters to give similar scores to different domains of a resident's performance even when the domains are clearly separate.^{4,18,25} Related to this is "grade inflation," which reflects ratings that are globally higher than a learner's performance warrants. Prior studies suggest that a committee process for making progress decisions may result in reduced grade inflation;²⁶ our study extends this literature by demonstrating a reduced halo effect through the use of a group assessment method. Interestingly, however, the average scores

Table 1. Individual Versus Group Assessment Item Mean Scores

Item	No.	Individual		Group		Sign rank p-value
		Mean	SD	Mean	SD	
Q1: Accuracy and completeness	148	3.92	0.48	3.93	0.73	0.67
Q2: Interpretation effectiveness	148	3.82	0.47	3.86	0.73	0.2
Q3: Select proper test	148	3.61	0.43	3.68	0.63	0.11
Q4: Illness management	148	3.88	0.45	3.96	0.65	0.03
Q5: Follow-up care	147	3.89	0.43	3.98	0.7	0.02
Q6: Patient retention	115	3.82	0.48	3.9	0.68	0.06
Q7: Commitment to education	147	3.87	0.41	4.02	0.64	0.01
Overall	148	3.83	0.38	3.92	0.51	0.0001

Table 2. Inter-Rater Reliability and Inter-item Correlation of Individual Assessments Alone and with the Addition of Group Assessments

Items	Individual assessments alone			Individual assessments with the addition of group assessments		
	No.	ICC [*]	95% CI [†]	No.	ICC	95% CI
Q1: Accuracy and completeness	644	0.661	0.577, 0.737	780	0.774	0.718, 0.824
Q2: Interpretation effectiveness	642	0.632	0.541, 0.714	778	0.763	0.703, 0.815
Q3: Select proper test	623	0.599	0.500, 0.688	761	0.711	0.639, 0.775
Q4: Illness management	639	0.585	0.482, 0.677	775	0.707	0.634, 0.772
Q5: Follow-up care	589	0.443	0.305, 0.566	724	0.588	0.486, 0.679
Q6: Patient retention	462	0.587	0.479, 0.683	567	0.694	0.615, 0.765
Q7: Commitment to education	580	0.545	0.433, 0.646	715	0.688	0.610, 0.757
Overall	644	0.749	0.686, 0.804	780	0.828	0.785, 0.866

	Individual assessments alone	Individual assessments with the addition of group assessments
Inter-item correlation	0.68	0.49

*ICC = Intraclass correlation coefficient, a measure of inter-rater reliability
[†]95% CI = 95% Confidence interval

across all performance items in our study increased with group assessments versus individual assessments, although by only a small amount. One explanation for this finding is that the group assessment improved the ability to target more specifically resident domains that were excellent, and those that required improvement, with the greater balance of emphasis being on domains of excellent performance.

We implemented group assessments in the outpatient setting to utilize the longitudinal relationship between our Continuity Clinic faculty and residents over their 3 years of training. However, most previous studies on the group assessment model occurred in the inpatient setting, where observations from faculty, senior residents, and junior residents were gathered for student assessment. Hemmer and colleagues have shown that formal evaluation sessions compared with individual

assessments better identified core clerkship students with marginal funds of knowledge and episodes of unprofessional behavior in the hospital setting,¹¹⁻¹³ and additionally served as case-based faculty development sessions.²⁷ The current study extends these findings to resident assessment in the context of the ACGME competencies and to the outpatient ambulatory clinic setting. This study also adds support to the important role formal evaluation sessions can serve in continuous faculty development. Schwind and colleagues found that discussion at a surgical resident evaluation committee, compared to individual assessments, identified an increased number of deficiencies across three performance domains.⁵ By comparison, our study provides evidence of significant differences in reliability and a halo effect between the two methods of assessment. Future studies should examine inpatient group assessment for resident

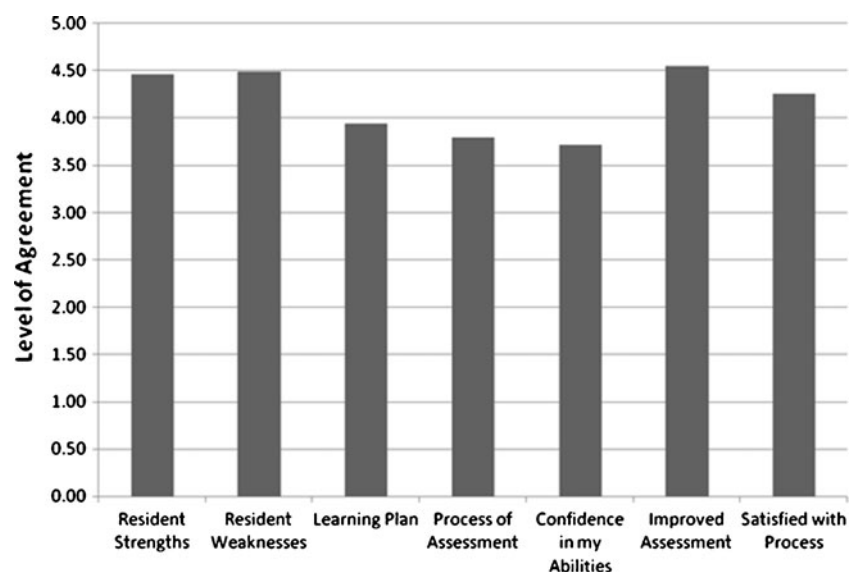


Figure 2. Faculty opinions regarding group assessments (n=87). Faculty members were asked immediately following each group assessment whether the discussion improved their understanding of resident strengths, weaknesses, and learning plans; improved their understanding of the process of assessment; improved their confidence in their own assessment abilities; and improved the overall process of assessment. They were then asked to rate their overall satisfaction with the process.

physicians, which could involve combined assessments from attending physicians, resident peers or senior medical residents, and allied health staff. Additionally, group assessments of senior residents might include the junior residents' feedback on teaching skills.

Group assessments in this study were remarkably inexpensive and efficient. Ninety-minute sessions were conducted quarterly, overlapping the lunch hour to minimize the impact on faculty's clinical productivity and other academic commitments. Lunch was provided, and meeting rooms with projectors—used to view photographs of residents and documented clinical observations—enhanced the discussions. Firm chiefs were given administrative time to compile information from the group assessment discussions. Finally, information discussed during the group assessments was obtained partly from individual continuity clinic faculty assessments (for example, direct observations of resident clinical performance), the completion of which is already a standing expectation for teaching faculty.

This study has several limitations. First, it was performed at a single institution, which may limit generalizability of the study findings. The group assessment model, however, has demonstrated feasibility at other institutions in clinical clerkships in both inpatient and outpatient settings.^{11,13,28} Given the relatively low costs of implementation and the adaptability of the group assessment model to different institutional environments and practice settings, we believe this model has universal appeal. Second, this study did not assess the impact of group assessments on improved resident outcomes. Nonetheless, we found that these assessments provide discriminating assessment data across domains of resident performance, which should allow more specific feedback to residents, thus enhancing their potential for improvement. Third, this study did not assess whether group dynamics influenced the groups' collective assessments. However, prior research on the effects of group dynamics on resident progress committee deliberations suggests that a controlled group decision-making process, such as used in our group assessment sessions, does not compromise the validity of assessment.²⁹ Fourth, although formal assessments of students have been found to help identify lapses in professionalism, we did not target the professionalism domain quantitatively by incorporating this into the assessment as a Likert-scaled item. However, we observed that narrative comments during the group assessments often provided information that could not be gleaned from the individual assessment scale data alone. Specifically, we found that issues pertaining to communication and professionalism often emerged through group discussion. This is similar to what we had found in previous qualitative studies at our institution, which showed that narrative comments provide, more than quantitative questionnaire data, valuable information regarding interpersonal dynamics.³⁰ Finally, faculty members' satisfaction and improved confidence with group assessment were surveyed with a non-validated instrument (so information regarding the properties of this instrument are unknown) and were not substantiated by objective measures. Future studies should assess whether these perceptions can be confirmed objectively.

In summary, this study demonstrates that integration of a group assessment in an internal medicine continuity clinic improved reliability and reduced range restriction (halo effect)

of resident assessment across multiple performance domains when compared to traditional individual faculty-on-resident assessment alone. This model, as one component of a global assessment system, should help graduate medical education programs achieve reliable and discriminating resident assessments despite increasing practice demands and faculty-resident discontinuity. Therefore, the group assessment might be considered as one of the solutions for addressing forthcoming changes in duty hour reform. Future research should assess the efficacy of this model for graduate medical education in hospital settings, investigate whether this model compared to traditional models of assessment leads to improved resident outcomes, and demonstrate whether use of this model leads to objective improvement in faculty assessment skills.

Acknowledgement: This project was supported in part through an Educational Innovations grant, Mayo Clinic Rochester.

Conflicts of Interest: None disclosed.

Corresponding Author: Matthew R. Thomas, MD; Division of Primary Care Internal Medicine, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA (e-mail: thomas.matthew@mayo.edu).

REFERENCES

- ACGME. Program Director Guide to the Common Requirements. 2011 [cited 2011 January 13]; Available from: <http://www.acgme.org/acWebsite/home/home.asp>.
- Chaudhry SI, Holmboe E, Beasley BW. The state of evaluation in internal medicine residency. *J Gen Intern Med.* 2008;23(7):1010–5.
- Epstein RM. Assessment in medical education. *N Engl J Med.* 2007;356(4):387–96.
- Gray JD. Global rating scales in residency education. *Acad Med.* 1996;71(1 Suppl):S55–63.
- Schwint CJ, Williams RG, Boehler ML, Dunnington GL. Do individual attendings' post-rotation performance ratings detect residents' clinical performance deficiencies? *Acad Med.* 2004;79(5):453–7.
- Levine HG, Yunker R, Bee D. Pediatric resident performance. The reliability and validity of rating forms. *Eval Health Prof.* 1986;9(1):62–74.
- Marienfeld RD, Reid JC. Six-year documentation of the easy grader in the medical clerkship setting. *J Med Educ.* 1984;59(7):589–91.
- Ryan JG, Mandel FS, Sama A, Ward MF. Reliability of faculty clinical evaluations of non-emergency medicine residents during emergency department rotations. *Acad Emerg Med.* 1996;3(12):1124–30.
- Thorndike EL. A constant error in psychological ratings. *J Appl Psychol.* 1920;4:25–9.
- McKinstry BH, Cameron HS, Elton RA, Riley SC. Leniency and halo effects in marking undergraduate short research projects. *BMC Med Educ.* 2004;4(1):28.
- Hemmer PA, Pangaro L. The effectiveness of formal evaluation sessions during clinical clerkships in better identifying students with marginal funds of knowledge. *Acad Med.* 1997;72(7):641–3.
- Hemmer PA, Grau T, Pangaro LN. Assessing the effectiveness of combining evaluation methods for the early identification of students with inadequate knowledge during a clerkship. *Med Teach.* 2001;23(6):580–4.
- Hemmer PA, Hawkins R, Jackson JL, Pangaro LN. Assessing how well three evaluation methods detect deficiencies in medical students' professionalism in two settings of an internal medicine clerkship. *Acad Med.* 2000;75(2):167–73.
- Albritton TA, Fincher RM, Work JA. Group evaluation of student performance in a clerkship. *Acad Med.* 1996;71(5):551–2.
- Beckman TJ, Mandrekar JN. The interpersonal, cognitive and efficiency domains of clinical teaching: construct validity of a multi-dimensional scale. *Med Educ.* 2005;39(12):1221–9.

16. **Beckman TJ, Mandrekar JN, Engstler GJ, Ficalora RD.** Determining reliability of clinical assessment scores in real time. *Teach Learn Med.* 2009;21(3):188–94.
17. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74.
18. **Risucci DA, Lutsky L, Rosati RJ, Tortolani AJ.** Reliability and accuracy of resident evaluations of surgical faculty. *Eval Health Prof.* 1992;15(3):313–24.
19. **Jacobs R, Kozlowski SW.** A closer look at halo error in performance ratings. *Academy of Management Journal.* 1985;28(1):pp.
20. **Murphy KR, Jako RA, Anhalt RL.** Nature and consequences of halo error: a critical analysis. *J Appl Psychol.* 1993;78(2):218–25.
21. **Beckman TJ.** Lessons learned from a peer review of bedside teaching. *Acad Med.* 2004;79(4):343–6.
22. **Iglehart JK.** Revisiting duty-hour limits—IOM recommendations for patient safety and resident education. *N Engl J Med.* 2008;359(25):2633–5.
23. **Holmboe ES, Rodak W, Mills G, McFarlane MJ, Schultz HJ.** Outcomes-based evaluation in resident education: creating systems and structured portfolios. *Am J Med.* 2006;119(8):708–14.
24. **Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski JJ.** Do global rating forms enable program directors to assess the ACGME competencies? *Acad Med.* 2004;79(6):549–56.
25. **Holmboe E, Hawkins RE.** *Practical Guide to the Evaluation of Clinical Competence.* Philadelphia: Mosby; 2008.
26. **Speer AJ, Solomon DJ, Ainsworth MA.** An innovative evaluation method in an internal medicine clerkship. *Acad Med.* 1996;71(1 Suppl):S76–8.
27. **Hemmer PA, Pangaro L.** Using formal evaluation sessions for case-based faculty development during clinical clerkships. *Acad Med.* 2000;75(12):1216–21.
28. **Battistone MJ, Milne C, Sande MA, Pangaro LN, Hemmer PA, Shomaker TS.** The feasibility and acceptability of implementing formal evaluation sessions and using descriptive vocabulary to assess student performance on a clinical clerkship. *Teach Learn Med.* 2002;14(1):5–10.
29. **Williams RG, Schwind CJ, Dunnington GL, Fortune J, Rogers D, Boehler M.** The effects of group dynamics on resident progress committee deliberations. *Teach Learn Med.* 2005;17(2):96–100.
30. **Kisiel JB, Bundrick JB, Beckman TJ.** Resident physicians' perspectives on effective outpatient teaching: a qualitative study. *Adv Health Sci Educ Theory Pract.* 2010;15(3):357–68.