

## ORIGINAL RESEARCH

## Implementing Peer Evaluation of Handoffs: Associations With Experience and Workload

Vineet M. Arora, MD, MAPP<sup>1,2\*</sup>, Elizabeth A. Greenstein, AB<sup>2</sup>, James N. Woodruff, MD<sup>1,2</sup>, Paul G. Staisiunas, BA<sup>1</sup>, Jeanne M. Faman, MD, MHPE<sup>1,2</sup>

<sup>1</sup>Section of General Internal Medicine, Department of Medicine, University of Chicago, Chicago, Illinois; <sup>2</sup>Pritzker School of Medicine, University of Chicago, Chicago, Illinois.

**BACKGROUND:** Although peer evaluation can be used to evaluate in-hospital handoffs, few studies have described using this strategy.

**OBJECTIVE:** Our objective was to assess feasibility of an online peer handoff evaluation and characterize performance over time among medical interns.

**DESIGN:** The design was a prospective cohort study.

**PATIENTS:** Subjects were medical interns from residency program rotating at 2 teaching hospitals.

**MEASUREMENTS:** Measurements were performance on an end-of-rotation evaluation of giving and receiving handoffs.

**RESULTS:** From July 2009 to March 2010, 31 interns completed 60% (172/288) of peer evaluations. Ratings were high across domains (mean, 8.3–8.6). In multivariate regression controlling for evaluator and evaluatee, statistically significant improvements over time were observed for 4 items compared to the first 3 months of the year: 1) communication

skills (season 2, +0.34 [95% confidence interval (CI), 0.08–0.60],  $P = 0.009$ ); 2) listening behavior (season 2, +0.29 [95% CI, 0.04–0.55],  $P = 0.025$ ); 3) accepting professional responsibility (season 3, +0.37 [95% CI, 0.08–0.65],  $P = 0.012$ ); and 4) accessing the system (season 2, +0.21 [95% CI, 0.03–0.39],  $P = 0.023$ ). Ratings were also significantly lower when interns were postcall in written sign-out quality (8.21 vs 8.39,  $P = 0.008$ ) and accepting feedback (8.25 vs 8.42,  $P = 0.006$ ). Ratings from a community hospital rotation, with a lower census than the teaching hospital, were significantly higher for overall performance and 7 of 12 domains ( $P < 0.05$  for all). Significant evaluator effects were observed.

**CONCLUSIONS:** Although there is evidence of leniency, peer evaluation of handoffs demonstrate increases over time and associations with workload such as postcall status. This suggests the importance of examining how workload impacts handoffs in the future. *Journal of Hospital Medicine* 2013;8:132–136. © 2013 Society of Hospital Medicine

The advent of restricted residency duty hours has thrust the safety risks of handoffs into the spotlight. More recently, the Accreditation Council of Graduate Medical Education (ACGME) has restricted hours even further to a maximum of 16 hours for first-year residents and up to 28 hours for residents beyond their first year.<sup>1</sup> Although the focus on these mandates has been scheduling and staffing in residency programs, another important area of attention is for handoff education and evaluation. The Common Program Requirements for the ACGME state that all residency programs should ensure that residents are “competent in handoff communications” and that programs should “monitor handoffs to ensure that they are safe.”<sup>2</sup> Moreover, recent efforts have defined milestones for handoffs, specifically that by 12 months, residents should be able to effectively communicate with other caregivers to main-

tain continuity during transitions of care.<sup>3</sup> Although more detailed handoff-specific milestones have to be flushed out, a need for evaluation instruments to assess milestones is critical. In addition, handoffs continue to represent a vulnerable time for patients in many specialties, such as surgery and pediatrics.<sup>4,5</sup>

Evaluating handoffs poses specific challenges for internal medicine residency programs because handoffs are often conducted on the fly or wherever convenient, and not always at a dedicated time and place.<sup>6</sup> Even when evaluations could be conducted at a dedicated time and place, program faculty and leadership may not be comfortable evaluating handoffs in real time due to lack of faculty development and recent experience with handoffs. Although supervising faculty may be in the most ideal position due to their intimate knowledge of the patient and their ability to evaluate the clinical judgment of trainees, they may face additional pressures of supervision and direct patient care that prevent their attendance at the time of the handoff. For these reasons, potential people to evaluate the quality of a resident handoff may be the peers to whom they frequently handoff. Because handoffs are also conceptualized as an interactive dialogue between sender and receiver, an ideal handoff performance evaluation would capture both of these roles.<sup>7</sup> For these reasons, peer evaluation may be a viable

\*Address for correspondence and reprint requests: Vineet Arora MD, University of Chicago, 5841 S Maryland Ave., MC 2007 AMB W216, Chicago, IL 60637; Tel.: (773) 702-8157; Fax: (773) 834-2238; E-mail: varora@medicine.bsdc.uchicago.edu

Additional Supporting Information may be found in the online version of this article.

Received: August 17, 2012; Revised: November 13, 2012; Accepted: November 15, 2012

2012 Society of Hospital Medicine DOI 10.1002/jhm.2002

Published online in Wiley Online Library (Wileyonlinelibrary.com).

modality to assist programs in evaluating handoffs. Peer evaluation has been shown to be an effective method of rating performance of medical students,<sup>8</sup> practicing physicians,<sup>9</sup> and residents.<sup>10</sup> Moreover, peer evaluation is now a required feature in assessing internal medicine resident performance.<sup>11</sup> Although enthusiasm for peer evaluation has grown in residency training, the use of it can still be limited by a variety of problems, such as reluctance to rate peers poorly, difficulty obtaining evaluations, and the utility of such evaluations. For these reasons, it is important to understand whether peer evaluation of handoffs is feasible. Therefore, the aim of this study was to assess feasibility of an online peer evaluation survey tool of handoffs in an internal medicine residency and to characterize performance over time as well and associations between workload and performance.

## METHODS

From July 2009 to March 2010, all interns on the general medicine inpatient service at 2 hospitals were asked to complete an end-of-month anonymous peer evaluation that included 14-items addressing all core competencies. The evaluation tool was administered electronically using New Innovations (New Innovations, Inc., Uniontown, OH). Interns signed out to each other in a “cross-cover circuit” that included 3 other interns on an every fourth night call cycle.<sup>12</sup> Call teams included 1 resident and 1 intern who worked from 7 AM on the on-call day to noon on the postcall day. Therefore, postcall interns were expected to hand off to the next on-call intern before noon. Although attendings and senior residents were not required to formally supervise the handoff, supervising senior residents were often present during postcall intern sign-out to facilitate departure of the team. When interns were not postcall, they were expected to sign out before they went to the clinic in the afternoon or when their foreseeable work was complete. The interns were provided with a 45-minute lecture on handoffs and introduced to the peer evaluation tool in July 2009 at an intern orientation. They were also prompted to complete the tool to the best of their ability after their general medicine rotation. We chose the general medicine rotation because each intern completed approximately 2 months of general medicine in their first year. This would provide ratings over time without overburdening interns to complete 3 additional evaluations after every inpatient rotation.

The peer evaluation was constructed to correspond to specific ACGME core competencies and was also linked to specific handoff behaviors that were known to be effective. The questions were adapted from prior items used in a validated direct-observation tool previously developed by the authors (the Handoff Clinical Evaluation Exercise), which was based on literature review as well as expert opinion.<sup>13,14</sup> For example, under the core competency of communication, interns were asked to

rate each other on communication skills using the anchors of “No questions, no acknowledgement of to do tasks, transfer of information face to face is not a priority” for low unsatisfactory (1) and “Appropriate use of questions, acknowledgement and read-back of to-do and priority tasks, face to face communication a priority” for high superior (9). Items that referred to behaviors related to both giving handoff and receiving handoff were used to capture the interactive dialogue between senders and receivers that characterize ideal handoffs. In addition, specific items referring to written sign-out and verbal sign-out were developed to capture the specific differences. For instance, for the patient care competency in written sign-out, low unsatisfactory (1) was defined as “Incomplete written content; to do’s omitted or requested with no rationale or plan, or with inadequate preparation (ie, request to transfuse but consent not obtained),” and high superior (9) was defined as “Content is complete with to do’s accompanied by clear plan of action and rationale.” Pilot testing with trainees was conducted, including residents not involved in the study and clinical students. The tool was also reviewed by the residency program leadership, and in an effort to standardize the reporting of the items with our other evaluation forms, each item was mapped to a core competency that it was most related to. Debriefing of the instrument experience following usage was performed with 3 residents who had an interest in medical education and handoff performance.

The tool was deployed to interns following a brief educational session for interns, in which the tool was previewed and reviewed. Interns were counseled to use the form as a global performance assessment over the course of the month, in contrast to an episodic evaluation. This would also avoid the use of “negative event bias” by raters, in which the rater allows a single negative event to influence the perception of the person’s performance, even long after the event has passed into history.

To analyze the data, descriptive statistics were used to summarize mean performance across domains. To assess whether intern performance improved over time, we split the academic year into 3 time periods of 3 months each, which we have used in earlier studies assessing intern experience.<sup>15</sup> Prior to analysis, postcall interns were identified by using the intern monthly call schedule located in the AMiON software program (Norwich, VT) to label the evaluation of the postcall intern. Then, all names were removed and replaced with a unique identifier for the evaluator and the evaluatee. In addition, each evaluation was also categorized as either having come from the main teaching hospital or the community hospital affiliate.

Multivariate random effects linear regression models, controlling for evaluator, evaluatee, and hospital, were used to assess the association between time (using indicator variables for season) and postcall status on intern performance. In addition, because of the skewness in

**TABLE 1.** Mean Intern Ratings on Sign-out Peer Evaluation by Item and Competency

ACGME Core Competency	Role	Items	Item	Mean	95% CI	Range	% Receiving 9 as Rating
Patient care	Sender	Written sign-out	Q1	8.34	8.25 to 8.48	6–9	53.2
	Sender	Updated content	Q2	8.35	8.22 to 8.47	5–9	54.4
	Receiver	Documentation of overnight events	Q6	8.41	8.30 to 8.52	6–9	56.3
Medical knowledge	Sender	Anticipatory guidance	Q3	8.40	8.28 to 8.51	6–9	56.3
	Receiver	Clinical decision making during cross-cover	Q7	8.45	8.35 to 8.55	6–9	56.0
Professionalism	Sender	Collegiality	Q4	8.60	8.51 to 8.68	6–9	65.7
	Receiver	Acknowledgement of professional responsibility	Q10	8.53	8.43 to 8.62	6–9	62.4
	Receiver	Timeliness/responsiveness	Q11	8.50	8.39 to 8.60	6–9	61.9
Interpersonal and communication skills	Receiver	Listening behavior when receiving sign-outs	Q8	8.52	8.42 to 8.62	6–9	63.6
	Receiver	Communication when receiving sign-out	Q9	8.52	8.43 to 8.62	6–9	63.0
Systems-based practice	Receiver	Resource use	Q12	8.45	8.35 to 8.55	6–9	55.6
Practice-based learning and improvement	Sender	Accepting of feedback	Q5	8.45	8.34 to 8.55	6–9	58.7
Overall	Both	Overall sign-out quality	Q13	8.44	8.34 to 8.54	6–9	55.3

NOTE: Abbreviations: ACGME, Accreditation Council of Graduate Medical Education; CI, confidence interval.

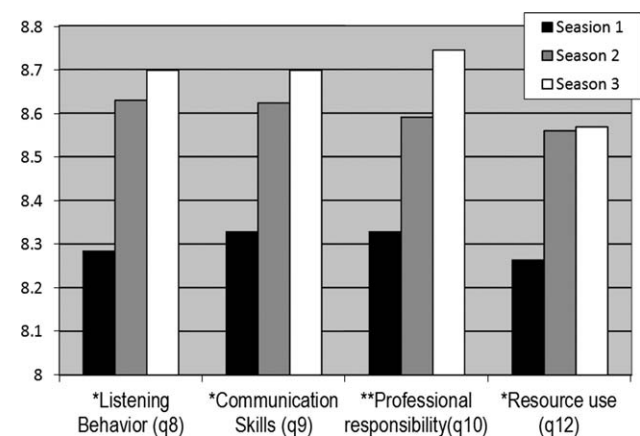
the ratings, we also undertook additional analysis by transforming our data into dichotomous variables reflecting superior performance. After conducting conditional ordinal logistic regression, the main findings did not change. We also investigated within-subject and between-subject variation using intraclass correlation coefficients. Within-subject intraclass correlation enabled assessment of inter-rater reliability. Between-subject intraclass correlation enabled the assessment of evaluator effects. Evaluator effects can encompass a variety of forms of rater bias such as leniency (in which evaluators tended to rate individuals uniformly positively), severity (rater tends to significantly avoid using positive ratings), or the halo effect (the individual being evaluated has 1 significantly positive attribute that overrides that which is being evaluated). All analyses were completed using STATA 10.0 (StataCorp, College Station, TX) with statistical significance defined as  $P < 0.05$ . This study was deemed to be exempt from institutional review board review after all data were deidentified prior to analysis.

## RESULTS

From July 2009 to March 2010, 31 interns (78%) returned 60% (172/288) of the peer evaluations they received. Almost all (39/40, 98%) interns were evaluated at least once with a median of 4 ratings per intern (range, 1–9). Thirty-five percent of ratings occurred when an intern was rotating at the community hospital. Ratings were very high on all domains (mean, 8.3–8.6). Overall sign-out performance was rated as 8.4 (95% confidence interval [CI], 8.3–8.5), with over 55% rating peers as 9 (maximal score). The lowest score given was 5. Individual items ranged from a low of 8.34 (95% CI, 8.21–8.47) for updating written sign-outs, to a high of 8.60 (95% CI, 8.50–8.69) for collegiality (Table 1). The internal consistency of the instrument was calculated using all items and was very high, with a Cronbach  $\alpha = 0.98$ .

Mean ratings for each item increased in season 2 and 3 and were statistically significant using a test for

trend across ordered groups. However, in multivariate regression models, improvements remained statistically significant for only 4 items (Figure 1): 1) communication skills, 2) listening behavior, 3) accepting professional responsibility, and 4) accessing the system (Table 2). Specifically, when compared to season 1, improvements in communication skill were seen in season 2 (+0.34 [95% CI, 0.08–0.60],  $P = 0.009$ ) and were sustained in season 3 (+0.34 [95% CI, 0.06–0.61],  $P = 0.018$ ). A similar pattern was observed for listening behavior, with improvement in ratings that were similar in magnitude with increasing intern experience (season 2, +0.29 [95% CI, 0.04–0.55],  $P = 0.025$  compared to season 1). Although accessing the system scores showed a similar pattern of improvement with an increase in season 2 compared to season 1, the magnitude of this change was smaller (season 2, +0.21 [95% CI, 0.03–0.39],  $P = 0.023$ ). Interestingly, improvements in accepting professional responsibility rose during season 2, but the difference did not reach statistical significance until season 3 (+0.37 [95% CI, 0.08–0.65],  $P = 0.012$  compared to season 1).



**FIG. 1.** Graph showing improvements over time in performance in domains of sign-out performance by season, where season 1 is July to September, season 2 is October to December, and season 3 is January to March. Results are obtained from random effects linear regression models controlling for evaluator, evaluatee, postcall status, and site (community vs tertiary).

**TABLE 2.** Increasing Scores on Peer Handoff Evaluation by Season

Predictor	Outcome β Coefficient (95% CI)				
	Communication Skills	Listening Behavior	Professional Responsibility	Accessing the System	Written Sign-out Quality
Season 1	Ref	Ref	Ref	Ref	Ref
Season 2	0.29 (0.04 to 0.55)*	0.34 (0.08 to 0.60)*	0.24 (−0.03 to 0.51)	0.21 (0.03 to 0.39)*	−0.05 (−0.25 to 0.15)
Season 3	0.29 (0.02 to 0.56)*	0.34 (0.06 to 0.61)*	0.37 (0.08 to 0.65)*	0.18 (0.01 to 0.36)*	0.08 (−0.13 to 0.30)
Community hospital	0.18 (−0.00 to 0.37)	0.23 (0.04 to 0.43)*	0.06 (−0.13 to 0.26)	0.13 (−0.00 to 0.25)	0.24 (0.08 to 0.39)*
Postcall	−0.10 (−0.25 to 0.05)	−0.04 (−0.21 to 0.13)	−0.02 (−0.18 to 0.13)	−0.05 (−0.16 to 0.05)	−0.18 (−0.31, −0.05)*
Constant	7.04 (6.51 to 7.58)	6.81 (6.23 to 7.38)	7.04 (6.50 to 7.60)	7.02 (6.59 to 7.45)	6.49 (6.04 to 6.94)

NOTE: Results are from multivariable linear regression models examining the association between season, community hospital, postcall status controlling for subject (evaluatee) random effects, and evaluator fixed effects (evaluator and evaluatee effects not shown). Abbreviations: CI, confidence interval. \* $P < 0.05$ .

In addition to increasing experience, postcall interns were rated significantly lower than nonpostcall interns in 2 items: 1) written sign-out quality (8.21 vs 8.39,  $P = 0.008$ ) and 2) accepting feedback (practice-based learning and improvement) (8.25 vs 8.42,  $P = 0.006$ ). Interestingly, when interns were at the community hospital general medicine rotation, where overall census was much lower than at the teaching hospital, peer ratings were significantly higher for overall handoff performance and 7 (written sign-out, update content, collegiality, accepting feedback, documentation of overnight events, clinical decision making during cross-cover, and listening behavior) of the remaining 12 specific handoff domains ( $P < 0.05$  for all, data not shown).

Last, significant evaluator effects were observed, which contributed to the variance in ratings given. For example, using intraclass correlation coefficients (ICC), we found that there was greater within-intern variation than between-intern variation, highlighting that evaluator scores tended to be strongly correlated with each other (eg, ICC overall performance = 0.64) and more so than scores of multiple evaluations of the same intern (eg, ICC overall performance = 0.18).

Because ratings of handoff performance were skewed, we also conducted a sensitivity analysis using ordinal logistic regression to ascertain if our findings remained significant. Using ordinal logistic regression models, significant improvements were seen in season 3 for 3 of the above-listed behaviors, specifically listening behavior, professional responsibility, and accessing the system. Although there was no improvement in communication, there was an improvement observed in collegiality scores that were significant in season 3.

## DISCUSSION

Using an end-of-rotation online peer assessment of handoff skills, it is feasible to obtain ratings of intern handoff performance from peers. Although there is evidence of rater bias toward leniency and low interrater reliability, peer ratings of intern performance did increase over time. In addition, peer ratings were lower for interns who were handing off their postcall service. Working on a rotation at a community affli-

ate with a lower census was associated with higher peer ratings of handoffs.

It is worth considering the mechanism of these findings. First, the leniency observed in peer ratings likely reflects peers unwilling to critique each other due to a desire for an esprit de corps among their classmates. The low intraclass correlation coefficient for ratings of the same intern highlight that peers do not easily converge on their ratings of the same intern. Nevertheless, the ratings on the peer evaluation did demonstrate improvements over time. This improvement could easily reflect on-the-job learning, as interns become more acquainted with their roles and efficient and competent in their tasks. Together, these data provide a foundation for developing milestone handoffs that reflect the natural progression of intern competence in handoffs. For example, communication appeared to improve at 3 months, whereas transfer of professional responsibility improved at 6 months after beginning internship. However, alternative explanations are also important to consider. Although it is easy and somewhat reassuring to assume that increases over time reflect a learning effect, it is also possible that interns are unwilling to critique their peers as familiarity with them increases.

There are several reasons why postcall interns could have been universally rated lower than nonpostcall interns. First, postcall interns likely had the sickest patients with the most to-do tasks or work associated with their sign-out because they were handing off newly admitted patients. Because the postcall sign-out is associated with the highest workload, it may be that interns perceive that a good handoff is “nothing to do,” and handoffs associated with more work are not highly rated. It is also important to note that postcall interns, who in this study were at the end of a 30-hour duty shift, were also most fatigued and overworked, which may have also affected the handoff, especially in the 2 domains of interest. Due to the time pressure to leave coupled with fatigue, they may have had less time to invest in written sign-out quality and may not have been receptive to feedback on their performance. Likewise, performance on handoffs was rated higher when at the community hospital, which could be due to several reasons. The most plausible

explanation is that the workload associated with that sign-out is less due to lower patient census and lower patient acuity. In the community hospital, fewer residents were also geographically co-located on a quieter ward and work room area, which may contribute to higher ratings across domains.

This study also has implications for future efforts to improve and evaluate handoff performance in residency trainees. For example, our findings suggest the importance of enhancing supervision and training for handoffs during high workload rotations or certain times of the year. In addition, evaluation systems for handoff performance that rely solely on peer evaluation will not likely yield an accurate picture of handoff performance, difficulty obtaining peer evaluations, the halo effect, and other forms of evaluator bias in ratings. Accurate handoff evaluation may require direct observation of verbal communication and faculty audit of written sign-outs.<sup>16,17</sup> Moreover, methods such as appreciative inquiry can help identify the peers with the best practices to emulate.<sup>18</sup> Future efforts to validate peer assessment of handoffs against these other assessment methods, such as direct observation by service attendings, are needed.

There are limitations to this study. First, although we have limited our findings to 1 residency program with 1 type of rotation, we have already expanded to a community residency program that used a float system and have disseminated our tool to several other institutions. In addition, we have a small number of participants, and our 60% return rate on monthly peer evaluations raises concerns of nonresponse bias. For example, a peer who perceived the handoff performance of an intern to be poor may be less likely to return the evaluation. Because our dataset has been deidentified per institutional review board request, we do not have any information to differentiate systematic reasons for not responding to the evaluation. Anecdotally, a critique of the tool is that it is lengthy, especially in light of the fact that 1 intern completes 3 additional handoff evaluations. It is worth understanding why the instrument had such a high internal consistency. Although the items were designed to address different competencies initially, peers may make a global assessment about someone's ability to perform a handoff and then fill out the evaluation accordingly. This speaks to the difficulty in evaluating the subcomponents of various actions related to the handoff. Because of the high internal consistency, we were able to shorten the survey to a 5-item instrument with a Cronbach  $\alpha$  of 0.93, which we are currently using in our program and have disseminated to other programs. Although it is currently unclear if the ratings of performance on the longer peer evaluation are valid, we are investigating concurrent validity of the shorter tool by comparing peer evaluations to other measures of handoff quality as part of our current work. Last, we are only able to test associations and not make causal inferences.

## CONCLUSION

Peer assessment of handoff skills is feasible via an electronic competency-based tool. Although there is evidence of score inflation, intern performance does increase over time and is associated with various aspects of workload, such as postcall status or working on a rotation at a community affiliate with a lower census. Together, these data can provide a foundation for developing milestones handoffs that reflect the natural progression of intern competence in handoffs.

## Acknowledgments

The authors thank the University of Chicago Medicine residents and chief residents, the members of the Curriculum and Housestaff Evaluation Committee, Tyrece Hunter and Amy Ice-Gibson, and Meryl Prochaska and Laura Ruth Venable for assistance with manuscript preparation.

Disclosures: This study was funded by the University of Chicago Department of Medicine Clinical Excellence and Medical Education Award and AHRQ R03 5R03HS018278-02 Development of and Validation of a Tool to Evaluate Hand-off Quality.

## References

1. Nasca TJ, Day SH, Amis ES Jr; the ACGME Duty Hour Task Force. The new recommendations on duty hours from the ACGME Task Force. *N Engl J Med*. 2010; 363.
2. Common program requirements. Available at: [http://acgme-2010standards.org/pdf/Common\\_Program\\_Requirements\\_07012011.pdf](http://acgme-2010standards.org/pdf/Common_Program_Requirements_07012011.pdf). Accessed December 10, 2012.
3. Green ML, Aagaard EM, Caverzagie KJ, et al. Charting the road to competence: developmental milestones for internal medicine residency training. *J Grad Med Educ*. 2009;1(1):5–20.
4. Greenberg CC, Regenbogen SE, Studdert DM, et al. Patterns of communication breakdowns resulting in injury to surgical patients. *J Am Coll Surg*. 2007;204(4):533–540.
5. McSweeney ME, Lightdale JR, Vinci RJ, Moses J. Patient handoffs: pediatric resident experiences and lessons learned. *Clin Pediatr (Phila)*. 2011;50(1):57–63.
6. Vidyarthi AR, Arora V, Schnipper JL, Wall SD, Wachter RM. Managing discontinuity in academic medical centers: strategies for a safe and effective resident sign-out. *J Hosp Med*. 2006;1(4):257–266.
7. Gibson SC, Ham JJ, Apker J, Mallak LA, Johnson NA. Communication, communication, communication: the art of the handoff. *Ann Emerg Med*. 2010;55(2):181–183.
8. Arnold L, Willouby L, Calkins V, Gammon L, Eberhardt G. Use of peer evaluation in the assessment of medical students. *J Med Educ*. 1981;56:35–42.
9. Ramsey PG, Wenrich MD, Carline JD, Inui TS, Larson EB, LoGerfo JP. Use of peer ratings to evaluate physician performance. *JAMA*. 1993;269:1655–1660.
10. Thomas PA, Gebo KA, Hellmann DB. A pilot study of peer review in residency training. *J Gen Intern Med*. 1999;14(9):551–554.
11. ACGME Program Requirements for Graduate Medical Education in Internal Medicine Effective July 1, 2009. Available at: [http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/140\\_internal\\_medicine\\_07012009.pdf](http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/140_internal_medicine_07012009.pdf). Accessed December 10, 2012.
12. Arora V, Dunphy C, Chang VY, Ahmad F, Humphrey HJ, Meltzer D. The effects of on-duty napping on intern sleep time and fatigue. *Ann Intern Med*. 2006;144(11):792–798.
13. Farnan JM, Paro JA, Rodriguez RM, et al. Hand-off education and evaluation: piloting the observed simulated hand-off experience (OSHE). *J Gen Intern Med*. 2010;25(2):129–134.
14. Horwitz LI, Dombroski J, Murphy TE, Farnan JM, Johnson JK, Arora VM. Validation of a handoff assessment tool: the Handoff CEX [published online ahead of print June 7, 2012]. *J Clin Nurs*. doi: 10.1111/j.1365-2702.2012.04131.x.
15. Arora VM, Georgitis E, Siddique J, et al. Association of workload of on-call medical interns with on-call sleep duration, shift duration, and participation in educational activities. *JAMA*. 2008;300(10):1146–1153.
16. Gakhar B, Spencer AL. Using direct observation, formal evaluation, and an interactive curriculum to improve the sign-out practices of internal medicine interns. *Acad Med*. 2010;85(7):1182–1188.
17. Bump GM, Bost JE, Buranosky R, Elnicki M. Faculty member review and feedback using a sign-out checklist: improving intern written sign-out. *Acad Med*. 2012;87(8):1125–1131.
18. Helms AS, Perez TE, Baltz J, et al. Use of an appreciative inquiry approach to improve resident sign-out in an era of multiple shift changes. *J Gen Intern Med*. 2012;27(3):287–291.