Description of a Developmental Criterion-Referenced Assessment for Promoting Competence in Internal Medicine Residents

ANDREW VARNEY, MD
CHRISTINE TODD, MD
SUSAN HINGLE, MD
MICHAEL CLARK

Abstract

Rationale End-of-rotation global evaluations can be subjective, produce inflated grades, lack interrater reliability, and offer information that lacks value. This article outlines the generation of a unique developmental criterion-referenced assessment that applies adult learning theory and the learner, manager, teacher model, and represents an innovative application to the American Board of Internal Medicine (ABIM) 9-point scale.

Intervention We describe the process used by Southern Illinois University School of Medicine to develop rotation-specific, criterion-based evaluation anchors that evolved into an effective faculty development exercise.

Results The intervention gave faculty a clearer understanding of the 6 Accreditation Council for Graduate Medical Education competencies, each rotation's educational goals, and how rotation design

affects meaningful work-based assessment. We also describe easily attainable successes in evaluation design and pitfalls that other institutions may be able to avoid. Shifting the evaluation emphasis on the residents' development of competence has made the expectations of rotation faculty more transparent, has facilitated conversations between program director and residents, and has improved the specificity of the tool for feedback. Our findings showed the new approach reduced grade inflation compared with the ABIM end-of-rotation global evaluation form.

Discussion We offer the new developmental criterion-referenced assessment as a unique application of the competencies to the ABIM 9-point scale as a transferable model for improving the validity and reliability of resident evaluations across graduate medical education programs.

Introduction

Statement of Purpose

Meaningful evaluation of all 6 Accreditation Council for Graduate Medical Education (ACGME) competencies has remained elusive for many residency programs. The July 2009 Program Requirements for Internal Medicine provide added impetus for improved formative and summative evaluations, including a directive to "define evaluation

expectations for each of the six ACGME competencies." Research shows that an evaluation with clear, observable criteria for each increment on the scale improves reliability,¹ and the ACGME suggests that criteria-based evaluations can serve as a motivation for residents to make improvements.² This article describes a process to apply criterion-referenced anchors for the assessment of internal medicine residents at Southern Illinois University School of Medicine.

The development of an end-of-rotation evaluation tool that uses competency-based criteria as the standards for resident assessment was prompted by our involvement in the ACGME Educational Innovations Project (EIP). Objectives included confronting grade inflation,^{3,4} poor interrater reliability^{3,5,6} and inconsistencies observed in faculty's understanding of the ACGME competencies and the American Board of Internal Medicine (ABIM) performance scale. Problems with assessment included differences among attending physicians in the elements that signified competence, and resident preoccupation with "looking good" to attending physicians instead of working on the attainment of competence. During program director feedback sessions, residents began to devalue the importance of faculty evaluations, and that they felt maligned when given ratings lower than their peers.

All authors are in the Department of Internal Medicine, Southern Illinois University School of Medicine: **Andrew Varney, MD**, is the Program Director and an Associate Professor; **Christine Todd, MD**, is the Associate Program Director and an Associate Professor; **Susan Hingle, MD**, is the Associate Program Director and an Associate Professor; and **Michael Clark** is the Educational Innovations Project Coordinator.

The authors received no outside funding for this work. To the authors' knowledge, no conflict of interest, financial or other, exists.

Special thanks to David Steward, MD, Department Chair, for his support in this project.

Thanks to Phillip Johnson, PhD, and Steve Sandstrom for their editorial review.

Thanks to Larry F. Hughes, PhD, for his statistical support.

Corresponding author: Andrew Varney, MD, PO Box 19636, Springfield, IL 62794-9636, 217.545.0193, avarney@siumed.edu

DOI: 10.4300/01.01.0012

Description of the Process

The Southern Illinois University (SIU) internal medicine residency has a decade of experience using the ABIM form for end-of-rotation evaluations. The form uses a 9-point Likert scale and criterion language for 1 and 9, but lacks descriptors for 2 through 8, causing frustration among faculty and residents. The problems we encountered with the prior evaluation system are similar to those experienced by other programs that use the ABIM format. Interrater reliability was low, and overall grades were inflated. Regardless of whether residents were in postgraduate year 1 or 3, they expected to receive an evaluation score of 8 or 9 on the 9-point scale, and lower scores were seen as punitive. We missed opportunities to provide valuable feedback to residents because of problems with the evaluation process, and we thought we could ameliorate this with a more criterion-based approach.

We revised the end-of-rotation form and developed a system to guide faculty and residents toward a core principle of competency. This resulted in a global evaluation formatted identically across all rotations but with unique anchor language for every rotation, with an innovative application of the 1 through 9 scale that clearly delineates the behaviors needed to reach each level of competency. Our Competency-Based Resident Assessment (CoBRA) offers a method for integration of the ACGME competencies into resident evaluation across specialties. Through development of a comparative database we were able to establish our historical baseline with the ABIM form, and we tracked individual faculty members' average resident evaluation score in each of the competencies for each level of training. This created a departmental average, allowing faculty members to compare their average to the departmental mean. Simply reporting this data back to the faculty did little to change evaluation behaviors, and this analysis confirmed some of our qualitative assessments previously described.

Before we began with the design of the new evaluation, a committee of physicians and staff reviewed the 6 competencies to determine the guidance available for writing specific evaluation criteria for each competency. The ACGME competencies gave us an ideal starting point, as each is composed of 3 subcompetencies or benchmarks. These not only lend themselves to anchor language, but also suggest a defined path to competency (TABLE 1). For example, the ACGME idealized patient care as "compassionate, appropriate, and effective," practice-based learning has 3 parts: "investigation and evaluation of their own patient care," "appraisal and assimilation of scientific evidence," and "improvements in patient care."⁷

Operationally, the set of subcompetencies for practice-based learning were translated as:

- Are residents willing to reveal what they do not know?

- Are they capable of searching out answers to clinical questions they develop?
- Are they able to apply the information in the correct clinical context?

We agreed to use the 3-part statements of the ACGME competencies on a 9-point scale with criterion-based language, and customize the form according to each rotation's needs. A benefit of asking faculty to develop the written criteria is that it challenged instructors to question what is "competent," state their expectations of the skills, knowledge, and attitudes that contribute to resident competency, and, in some cases, redesign the learning environment to facilitate the direct observation of these behaviors.

Evaluation Theory and Constructs

Our goal was to create a process consistent with evaluation models proposed by Pangaro,⁸ Miller,⁹ and Dreyfus¹⁰ and use the ABIM 1 to 9 scale. We decided to overlay onto the 1 to 9 scale the "learner, manager, teacher" (LMT) constructs and applied George Miller's "pyramid of learning." The LMT construct translated the criterion-referenced performance rating into functions that were broadly accepted by the residents and faculty.¹¹ Miller⁹ promoted a framework for assessing clinical competence with either standardized patients or in work situations, and proposed 4 levels: knows, knows how, shows how, and does (FIGURE 1). At SIU, faculty gives a 1 to 4 rating when residents are unable to meet the expectations of faculty. *Knowing*, but not *knowing how* is a critical distinction as residents transition from learner to manager. The ABIM form equates a performance rating of 1 to 3 as "may endanger patient safety," and a 4 as "marginal." Our interpretation of a learner as not yet ready to perform the duties and responsibilities of a resident/manager is internally consistent with this. We use a 5 to 7 rating when the faculty recognizes the resident as an effective manager who meets the faculty's expectations across the competencies by the end of the rotation. We reserve an 8 rating for residents who show an ability to teach and lead in the subcompetencies. Only the Clinical Competency Committee can award a rating of a 9. It requires a score of 8 on all competencies and publications or another meaningful contribution to the goals of the EIP.

FIGURE 2 shows the scale as it appears on the end-of-the-month global evaluation forms. No numbers are used; a checkbox indicates only criteria mastered and those not yet met. Faculty writes 2 criteria-based statements for each benchmark or subcompetency. The first criterion describes the expectation for residents at the beginning of the rotation, and the second describes the expectation for residents as they leave the rotation. Faculty evaluates residents only at the end of the rotation, placing the resident along a developmental continuum and

TABLE 1 ACGME COMPETENCIES WITH BENCHMARKS

For the development of our evaluation scale, we used the 3-part design of the competencies. The first sentence is the exact wording of the 2007 ACGME Competencies for Internal Medicine. We added underlining for emphasis. The following is a breakout of the 3 composite parts, or benchmarks, as we interpret the definitions in the 2007 ACGME Outcome Project as applicable.

WHAT YOU DO^a

Patient Care that is compassionate, appropriate, and effective for the treatment of health problems and promotion of health.

WHAT YOU KNOW^a

Medical Knowledge about established and evolving biomedical, clinical, and cognate (eg, epidemiological and social-behavioral) sciences and the application of this knowledge to improvement in patient care.

HOW YOU GET BETTER^a

Practice-Based Learning and Improvement that involves investigation and evaluation of their own patient care, appraisal and assimilation of scientific evidence, and improvements in patient care.

1. **Self-Assessment and Reflection** that involves investigation and evaluation of their own practice
2. **Self-Directed Learning** that involves investigation, evaluation, appraisal, and assimilation of scientific evidence into the practice (eg, includes the willingness to seek out answers to what is not known and the appropriate application of new information to improve patient care.)
3. **Practice Modification** that involves investigation and evaluation of improvement in patient care (includes patient safety, quality improvement and error reduction activities/awareness, and teaching.)

HOW YOU INTERACT WITH OTHERS^a

Interpersonal and **Communication Skills** that result in effective information exchange and teaming with patients, their families, and other health professionals.

1. **Communication skills** that result in effective information exchange (eg, listening, nonverbal, explanatory, questioning, and writing) with patients, families and other professionals
2. **Interpersonal skills** that result in shared decision-making to create a therapeutic and ethically sound relationship with patients, families, or surrogates (using communication skills effectively to create a productive medical and therapeutic interaction between doctor, patient and others)
3. **Teamwork** Using communication and interpersonal skills to promote working as a leader of, or member of, a team with patients, families, and other health professionals

HOW YOU ACT^a

Professionalism, as manifested through a commitment to carrying out professional responsibilities, adherence to ethical principles, and sensitivity to a diverse patient population.

HOW YOU WORK WITHIN THE SYSTEM^a

Systems-Based Practice, as manifested by actions that demonstrate an awareness and responsiveness to the larger context and system of health care and the ability to effectively call on system resources to provide care that is optimal value.

1. **Access**: as manifested by actions that demonstrate an awareness of and responsiveness to the larger context of health care (understand the barrier to, and the access to care in community, social services, and self-help such as homeless shelters, AA, Visiting Nurses, Veterans Administration, free clinics, pharmacy program assistance, etc.)
2. **Cost and Quality**: As manifested by actions that demonstrate and awareness of the system of health care to provide care that is of optimal value (eg cost effective quality care) for both individual patients and panels.
3. **Networking** and Partnering with other Teams for Affordable, Safe, Patient Centered Care: as manifested by actions that demonstrate an awareness and the ability to call on system resources (the ability to network/partner with teams that use available resources) to provide care that is of optimal value (cost conscious/cost effective for both the system and the patient)

^a Headings taken from Linda Tomasa and the Educational Outcomes Service Group (EOSG).

No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, web distribution or information storage and retrieval systems—without the written permission of SIU.

Copyright © 2008, All Rights Reserved.

documenting his/her progress from learner to manager to teacher/leader. Because a rating of 9, “scholar,” is awarded only by the Clinical Competency Committee, this score was removed from evaluation form.

The following paragraphs describe the levels mentioned on FIGURE 2.

Level I–Learner. The baseline or minimum knowledge that faculty expect of residents before they start a rotation (1–4).

Level II–Manager. The knowledge that faculty expects residents to demonstrate when they finish the rotation.

These are an extension of Level I but contain more advanced performance criteria (5–7).

Level III–Teacher/Leader. Mastery of all parts of Levels I and II, and the ability to disseminate to others an understanding of the multifaceted correct methods to deliver safe and patient-centered medical care (8).

Level IV–Scholar (not shown on the evaluation form). The promotion of the goals of the EIP and/or publishing in peer-reviewed journals (9).

The LMT overlay allowed us to define more clearly how this scale measures competency. Because each step on the

FIGURE 2

END-OF-MONTH GLOBAL EVALUATION SCALE FOR RESIDENTS, SOUTHERN ILLINOIS UNIVERSITY SCHOOL OF MEDICINE

Fails all criteria	Achieves one Level I criteria	Achieves two Level I criteria	Achieves all Level I criteria	Achieves all Level I and one Level II criteria	Achieves all Level I and two Level II criteria	Achieves all Level I and Level II criteria	Achieves all Level I and II & teaches
☐	☐	☐	☐	☐	☐	☐	☐
LEARNER				MANAGER		TEACHER/LEADER	

scale represents mastery of the materials representing all prior numbers (a Guttman scale), we could use this method of evaluation to foster constructive *developmental* assessment sessions. In a Guttman scale, items have a cumulative property and are arranged from least to most extreme. If a resident fails an item, he/she cannot attain the next higher level until mastering that item.¹² Lou Pangaro⁸ incorporated a Guttman scale in his 4-stage reporter-interpreter, manager, and educator (RIME) model, recommending that residents be elevated to the next level on the scale only when the subordinate skills have been mastered and demonstrated.

We eliminated the numbers on the evaluation form to focus residents' attention on personal growth in the field of medicine, rather than the achievement of top scores. The numbers have been retained for our comparative database, for reporting to the ABIM. A goal for our evaluation sessions is for faculty and residents to discuss specific, concrete issues related to the anchoring language to improve the accuracy of the evaluation and the quality of the feedback.

Discussion of Anchoring Language by Division

Instead of using a department-wide evaluation, we asked each division sponsoring a rotation to prepare anchoring language to evaluate each competency. We did this for several reasons. First, we wanted to recognize that different groups of physicians could have different views on behaviors that demonstrate competency in the specialty and wanted anchoring language that reflected each rotation's particular interpretation of the ACGME competencies to assist with orientation ("this is what we expect") and for use in constructive feedback. Second, we recognized the potential for faculty development around the competencies, particularly practice-based learning and systems-based practice. We wanted faculty to work collaboratively to set subcompetency expectations, and wanted their buy-in for the change in evaluation methodology.

We supplied faculty with tools, including the annotated descriptions of the 6 competencies (TABLE 1), Bloom's taxonomy of high-order thinking skills, that instructors traditionally use to write evaluations, and examples of observable criteria from other residency programs' evaluations listed under each of the 3 parts of the 6 competencies.¹³ This provided faculty with specific, objective definitions. The overall approach focused on measuring the resident against the scale.¹⁴

The next step involved integration of this plan into our database using a commercial evaluation software system. Multiple formats were developed and explored, from yes/no statements to incorporating the scale into the assessment of each competency. All evaluation forms used in our program were modified to reflect the 1 to 8 scale but only the rotational end-of-month forms have been anchored around a Guttman scale. Expectations for resident performance that includes but goes beyond rotation performance are termed

milestones of participation (TABLE 2). Each area of nonparticipation for a given resident will reduce the score in professionalism reported to the ABIM by one-half of a point.

We discovered that faculty required a great deal of support in writing anchors for the subcompetencies. The educational bonding generated by faculty during an inclusive review process was absolutely crucial to a good end product, and problems arose when divisions asked a single faculty member to write the anchors. We introduced the new evaluation to residents in a series of all-resident meetings, emphasizing the ability of this approach to show residents the specific skills or elements of performance they needed to address. Residents quickly accepted the idea that an intern might begin with a rating of 3—a score previously understood as disastrous at any stage of training—and the process focused on whether he/she moved up the scale during training. Residents noted that the clear expectations provided through the new process encouraged them to "show" attending physicians their skill level, a definite boost for the level of enthusiasm in our program.

Early Experience With Implementation: Perils and Promise

After implementing the new evaluation, it became clear that problems existed in nearly all rotations. A common issue involved ensuring that faculty had sufficient time to directly observe residents to assess their performance. The literature provides ample evidence that direct observation of resident performance is essential for valid and reliable evaluation,¹⁵ but the working lives of faculty and residents have been changed irrevocably by adoption of the ACGME duty hour standards and caps on admissions and patient census. Faculty across the country spend too little time observing residents as they interact with patients.¹⁶ To address this issue, some divisions restructured their learning environment; others revised their competency expectations, a process that we supported—as long as it was done with academic integrity. Few rotations successfully implemented the first iteration of the evaluation without the need for changes.

One solution for restructuring came from our hospitalist rotation: a faculty evaluator who was not on service was placed on hospital patient care units as a resident evaluator. The quality of education and evaluation improved, yet this approach likely is not possible on every rotation. Another solution adopted by many rotations to gain more individual input into the evaluations included meeting at the end of each rotation and filling in a group evaluation based on documentation completed earlier. A group evaluation has the advantage of providing a broader perspective on performance with less subjectivity than traditional end-of-rotation evaluations completed by individuals. Group meetings also allow for ongoing "calibration" of divisional faculty as they use the methodology and provide a work-around when individual attending physicians spend too

TABLE 2

SOUTHERN ILLINOIS UNIVERSITY SCHOOL OF MEDICINE MILESTONES OF PARTICIPATION

PGY-1	PGY-2	PGY-3
1. Successful completion of assigned rotations	1. Successful completion of assigned rotations	1. Successful completion of assigned rotations
Direct observation by faculty summarized in end-of-rotation evaluation	Direct observation by faculty summarized in end-of-rotation evaluation	Direct observation by faculty summarized in end-of-rotation evaluation
2. Attend 60% of conferences and complete journal club requirements	2. Attend 60% of conferences and complete journal club requirements	2. Attend 60% of conferences and completes journal club requirement
Observed and recorded by staff	Observed and recorded by staff	Observed and recorded by staff
3. Complete 70% ambulatory conferences	3. Complete 70% ambulatory conferences	3. Complete 70% ambulatory conferences
Observed and recorded by staff	Observed and recorded by staff	Observed and recorded by staff
4. Complete mini-CEXs	4. Complete mini-CEXs	4. Complete mini-CEXs
A. Focused history	A. Pain assessment	A. Complete H & P
B. Sexual history	B. Cardiovascular exam	B. EKG interpretation
C. Respiratory exam	C. Abdominal exam	C. Acid-base interpretation
D. Funduscopic exam	D. Neuro exam	D. Breaking bad news
	E. Musculoskeletal exam	
	F. Ventilator management	
	G. Thyroid exam	
	H. General counseling skills	
5. Complete 70% peer evaluations	5. Complete 70% peer evaluations	5. Complete 70% peer evaluations
6. Complete 70% student evaluations	6. Complete 70% student evaluations	6. Complete 70% student evaluations
7. Complete 70% faculty evaluations	7. Complete 70% faculty evaluations	7. Complete 70% faculty evaluations
8. Mandatory participation (60%) in board review questions if ITE falls below 30th percentile	8. Mandatory participation (60%) in board review questions if ITE falls below 30th percentile	8. Mandatory participation (60%) in board review questions if ITE falls below 30th percentile
	9. Complete resident-led QI project	9. Complete resident-led QI project
	10. Clinical vignette	10. Clinical vignette
	11. Pass Step 3 USMLE	11. Complete geriatric experience
		12. Attend 130 continuity clinics

Abbreviations: CEX, clinical evaluation exercise; ECG, electrocardiogram; H & P, history and physical examination; ITE, in-training examination; PGY, postgraduate year; QI, quality improvement; USMLE, United States Medical Licensing Examination.

little time with residents to make an evaluation of all 6 competencies. Fears of one strong personality dominating the discussion or opinion of the resident are minimized when faculty follow guidelines.¹⁷ The group evaluation also solved the problem that a single evaluator was “unable to assess” a dimension of performance because it was likely observed by another faculty member.

The commitment required for direct observation raised questions about what was being measured: clinical performance or clinical competency. We reconciled this question by noting it is the sum of all performances

measured against the standards of competency that determined a resident’s functional level. Divisions responded by creating an end-of-rotation evaluation. At the departmental level, we tracked every faculty member’s participation as part of our faculty evaluation dashboard, with the potential to use this information for promotion and tenure.

We have begun to gather data to determine if the new process is indeed more effective than the previous evaluation model. Initial data stem from one categorical residency program spanning 19 months of development and

TABLE 3
INTERNS (POSTGRADUATE YEAR-1) RECEIVING A SCORE OF 8

Competency	% of Interns Awarded an 8 Rating With ABIM Form	% of Interns Awarded an 8 Rating With CoBRA	Chi Square	P Value
Patient care	23 (146 of 636)	9.8 (53 of 540)	37.940	≤.001
Medical knowledge	19 (120 of 636)	11.3 (54 of 479)	13.521	≤.001
Practice-based learning	21.2 (123 of 580)	10.2 (49 of 484)	26.281	≤.001
Interpersonal communication	25.7 (123 of 580)	10.4 (50 of 479)	41.613	≤.001
Professionalism	33.6 (195 of 580)	12.1 (58 of 479)	69.400	≤.001
System-based practice	17.1 (109 of 636)	10 (48 of 480)	13.042	≤.001
Overall average	23.1 (842 of 3 648)	10.6 (312 of 2 941)	187.440	≤.001

Abbreviations: ABIM, American Board of Internal Medicine; CoBRA, competency-based resident assessment.

implementation of CoBRA and shows statistically significant improvements in evaluation. TABLE 3 shows how the new evaluation form reduced the number of 8 ratings awarded by faculty to interns, even when correcting for the number of tests done (Bonferroni correction).

TABLE 4 shows how the new method of evaluation increased the number of 3 ratings earned by residents by competency across the 3 years of training. Before the adjustment for the number of tests done (Bonferroni correction), all increases were statistically significant. However, after application of the Bonferroni adjustment, increases to the number of 3 ratings awarded in interpersonal and communication skills and system-based practice failed to retain statistical significance.

Early Experience With the Feedback Tool

Our early experience with feedback to residents and faculty has yielded positive results. The faculty who took part in the collaborative effort to create the criteria offered anecdotal evidence of significant improvement of feedback to residents. Faculty members think they are more effective in communicating strengths and weaknesses to residents. Residents now can review the assessment tool prior to the rotation, and think that faculty expectations are clearly delineated.

By entering annual data into a comparative database, program directors can use CoBRA to determine the mean score for a competency among all residents at any level of training. Each resident can then compare his/her mean score with the resident average. These data indicate much more than identifying “hawks” (ie, strict raters) and “doves” (ie, lenient raters). It exposes evaluators who routinely skip evaluating certain competencies, such as practice-based learning and improvement (PBLI) or system-based practice, and reveals weaknesses in the curriculum when all or most residents fail certain criteria. We are encouraged by the

statistically significant trends observed in our comparison of CoBRA with the ABIM form for end-of-rotation evaluations. Grade inflation has been positively affected, as the number of interns receiving a score of 8 has decreased, and the number of 3 ratings has increased across all levels of training.

Across the 6 ACGME competencies, we have created a list of observable behaviors and chose to incorporate the departmental values of teaching excellence and scholarship into the scale. We are optimistic this methodology will provide important feedback tailored more specifically to residents’ month-by-month progress by reducing the focus on numerical scores and the faculty propensity to inflate them.

Greater alignment between the curriculum guides and evaluation is another benefit derived from the creation of division-specific criteria. We urged divisions to use their curriculum guides in writing their criteria. Where curriculum guides were weak, faculty strengthened them before beginning the development of the criteria. In divisions with strong, clear curriculum guides, alignment with the criteria was easier and the transition nearly seamless. Another positive aspect of CoBRA is ongoing faculty development. Requiring faculty to write criteria based on the competencies forces the faculty to face its level of comfort with all of them.

Conclusion

Ownership of the evaluation criteria by all faculty members has been a key element of the success we have experienced to date. The project has proven to be a labor-intensive undertaking, but the ongoing faculty development, the focus on evaluation as a part of teaching, and the clarifying effects on goals and objectives within our department are positive returns for the time investment.

TABLE 4 RESIDENTS RECEIVING A SCORE OF 3^a

Competency	% of Residents Awarded a 3 Rating With ABIM Form	% of Residents Awarded a 3 Rating With CoBRA	Chi Square Over 3 Years	P Value Over 3 Years
Patient care			7.610	.006
PGY-1	0 (0 of 636)	0.56 (3 of 540)		
PGY-2	0 (0 of 641)	0.47 (2 of 429)		
PGY-3	0.20 (1 of 507)	0.25 (1 of 404)		
Medical knowledge			9.415	.002
PGY-1	0 (0 of 636)	1.0 (5 of 479)		
PGY-2	0.16 (1 of 641)	0.95 (4 of 420)		
PGY-3	0.20 (1 of 507)	0 (0 of 372)		
PBLI			20.822	.001
PGY-1	0 (0 of 580)	2.48 (12 of 484)		
PGY-2	0 (0 of 581)	0.89 (4 of 448)		
PGY-3	0.21 (1 of 476)	0.27 (1 of 373)		
Communication			6.096	.0135
PGY-1	0.69 (4 of 580)	1.46 (7 of 479)		
PGY-2	0 (0 of 581)	0.95 (4 of 420)		
PGY-3	0.21 (1 of 476)	0.27 (1 of 372)		
Professionalism			7.524	.006
PGY-1	0 (0 of 580)	0.63 (3 of 479)		
PGY-2	0 (0 of 581)	0.46 (2 of 420)		
PGY-3	0.20 (1 of 476)	0.27 (1 of 372)		
System-based practice			4.822	.028
PGY-1	0.31 (2 of 636)	1.04 (7 of 480)		
PGY-2	0.16 (1 of 638)	0.46 (2 of 431)		
PGY-3	0.39 (2 of 507)	0.27 (1 of 374)		

Abbreviations: ABIM, American Board of Internal Medicine; CoBRA, competency-based resident assessment; PBLI, practice-based learning and improvement; PGY, postgraduate year.

^a It may also be noted that PGY-3 residents, on average, were awarded fewer 3 ratings than PGY-2 residents (5 of 2 219 vs 18 of 2 340, $\chi^2 = 6.71$, $P = .010$) and far fewer than PGY-1 residents (5 of 2 219 vs 37 of 2 923, $\chi^2 = 17.21$, $P < .001$). The proportion of PGY-1 and PGY-2 residents that were awarded 3 ratings was not statistically different, although fewer were awarded to PGY-2 residents than PGY-1 residents.

Our approach has limitations. A single measurement tool cannot successfully evaluate residents in a comprehensive or perfect way. Research has demonstrated that using multiple evaluation methods is most efficacious.¹⁸ The SIU Internal Medicine Residency has a multisource evaluation system that informs the semiannual feedback sessions of the program director and, later, the resident portfolio; the CoBRAs are just one part of a system of resident assessment. Our findings also are limited to a single department's data and culture. We look forward to the

opportunity to collaborate with members of the learning community created around the EIP programs on a broader test of CoBRA. Plans call for implementation of CoBRA across several EIP programs on a hospital-based rotation. We believe the process described here shows great promise as a "transportable" component to other graduate medical education programs and will foster greater transparency in global evaluations. It is our hope that the graduate medical education community will explore, refine, and adopt the CoBRA scale into the fabric of its evaluation process.

References

- 1 Scholfield P. *Quantifying Language*. Philadelphia, PA: Clevedon, Avon, Multilingual Matters; 1995:209.
- 2 Leach D. Unlearning: it is time. *ACGME Bulletin*, April 2005;2.
- 3 Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287:226–235.
- 4 Swing S. Assessing the ACGME general competencies: general considerations and assessment methods. *Acad Emerg Med*. 2002;9:1278–1288.
- 5 Warf BC, Donnelly MB, Schwartz MW, et al. The relative contributions of interpersonal and specific clinical skills to the perception of global clinical competence. *J Surg Res*. 1999;86:17–24.
- 6 Cohen R, Rothman AI, Poldre R, et al. Validity and generalizability of global ratings in an objective structured clinical examination. *Acad Med*. 1991;66:545–548.
- 7 Accreditation Council for Graduate Medical Education. Competency language, common program requirements. Effective July 1, 2007. Available at: <http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf>. Accessed July 17, 2009.
- 8 Pangaro L. A new vocabulary and other innovation for improving descriptive in-training evaluations. *Acad Med*. 1999;74(11):1203–1207.
- 9 Miller G. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(suppl): S63–S67.
- 10 Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. *Health Aff (Millwood)*. 2002;21(5):103–111.
- 11 Rosenblum MJ, Borden SH, McArdle P, et al, Baystate Medical Center EIP Workgroup. The Baystate Manager Model. *Acad Intern Med Insight*. 2007;5(2):18.
- 12 Blalock, HM. *Social Statistics*. New York, NY: McGraw-Hill; 1979:22–23.
- 13 Huddle TS, Heudebert GR. Taking apart the art: the risk of anatomizing clinical competence. *Acad Med*. 2007;82(6):536–541.
- 14 Penn State Learning Design Community Hub; Bloom's Taxonomy; Key verbs <http://ets.fit.psu.edu/learningdesign/objectives/bloom>.
- 15 Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med*. 2004;140(11):874–881.
- 16 Holmboe ES, Fiebach NF, Galaty L, Huot S. The effectiveness of a focused educational intervention on resident evaluations from faculty: a randomized controlled trial. *J Gen Intern Med*. 2001;16:427–434.
- 17 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.
- 18 Wass V, van der Vleuten CPM, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357:945–949.