Are the Communication and Professionalism Competencies the New Critical Values in a Resident's Global Evaluation Process?

Mounir J. Haurani, I. Rubinfeld,* S. Rao,* J. Beaubien,† J. L. Musial,* A. Parker,* C. Reickert,* A. Raafat,* and A. Shepard**

*Department of Surgery, Henry Ford Hospital, Detroit, Michigan; and †Human-Systems Integration Division, Aptima, Inc., Woburn, Massachusetts

BACKGROUND: The ACGME requires the assessment of resident competency in 6 domains. Global evaluations covering all 6 competencies are routinely used. Evaluators may be overly influenced by resident affability and availability, thereby resulting in a halo effect. We hypothesized that the Interpersonal Skills and Communications (ICS) and Professionalism (PR) competencies would unduly influence other competency scores.

METHODS: General surgery resident evaluations are performed by staff and peers on a rotational basis using competency-based questions. Each question is scored using a 5-point Likert scale. Mean individual composite scores for each competency were calculated and then correlated with other mean composite competency scores. Data from patient evaluations were similarly analyzed. A final correlation of competency scores to ABSITE scores, as an objective, standardized measure of a specific competency, Medical knowledge (MK) was also performed.

RESULTS: Results were available for 37 residents (PGY 1-5). There was a significant association between ICS scores and higher scores in MK ($r = 0.52$, $p = 0.004$), PR ($r = 0.826$, $p < 0.0001$) and patient care (PC) ($r = 0.619$, $p < 0.0001$). No correlation, however, was found between patient evaluations of residents and their faculty/peer-based ICS scores. We found no association between ICS scores and improved patient evaluations. Lastly, we found no association between ICS or MK scores and ABSITE scores.

CONCLUSIONS: It was difficult to ascertain whether residents with better ICS scores had higher PR, PC, and MK scores because of the halo effect, improper completion of evaluations, or whether those residents were truly performing better clinically. External measures of resident performance did not correlate with faculty/peer evaluations of ICS and PR. Residency programs should consider adopting a more standardized way to objectively evaluate residents. (J Surg 64:351-356. © 2007 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: resident training, evaluation, competencies, resident evaluation, communications competency, professionalism, halo effect, global evaluations

COMPETENCY: Patient Care, Medical Knowledge, Professionalism, Interpersonal and Communication Skills, Practice Based Learning and Improvement

The Accreditation Council of Graduate Medical Education (ACGME) first created and then mandated use of the 6 competencies through its Outcomes Project.[1] This initiative requires the assessment of resident competency in 6 domains with little guidance or recommendations for standardization of methods. Although the ACGME has clearly stated what the expectations are, it has not given surgical education programs more than suggestions for measuring progress and proficiency in these domains. Yet, the current expectation is that programs will not only measure resident performance in terms of these competencies, but also document steady progression and improvement.

Many strategies have been used to increase the reliability and utility of the resident evaluation process. A variety of methods have been described, including standardized examinations, oral examinations, objective standardized clinical examinations (OSCEs), and skills laboratories. The global evaluation has been a particularly popular assessment tool used nearly universally by a wide variety of residency programs. With this tool residents are reviewed at intervals (eg, monthly or by rotation) using a set of questions addressing specific competency-based behaviors and/or proficiencies. Responses are frequently scored on a 5-point or 10-point Likert scale[2] to attempt standardization of results between subjects and evaluators.[3] To accurately

Correspondence: Inquiries to Mounir J. Haurani, General Surgery, Henry Ford Hospital, 2799 West Grand Blvd, Detroit, MI 48202; e-mail: mhauran1@hfhs.org

evaluate a resident in these areas requires that reviewers be cognizant of the various subcomponents of each competency. This requirement can be met by fully educating evaluators about the competencies and/or by designing an assessment tool that encompasses all subcomponents of the competencies and clearly defines them. Faculty training is also necessary to ensure that responses are calibrated and reproducible between all faculty members; such training helps assure an acceptable inter-rater reliability necessary for evaluation validity.[4-6] In reality, the logistics related to faculty training (ie, the educational process necessary to create and maintain such calibration) is a major barrier to its appropriate and widespread adoption.

Although valuable, faculty evaluations may not reveal a complete picture of resident performance. To provide more comprehensive assessment, many authorities have adapted a multisource (360°) feedback evaluation, where all members of the health-care team evaluate resident performance. The concept of 360° evaluations was popularized in the business sector and has recently been accepted in medical education.[7-9] Residents are evaluated by other residents at different levels of training, nurses, social workers, and even patients. The 360° evaluation process may be more consistent than other types of evaluations and is thought to provide better insight into the "softer" competencies such as professionalism (PR) and interpersonal and communication skills (ICS), which include interactions with peers, patients, and other members of the health-care team.[10,11] Despite the added effort and paperwork involved in compiling 360° evaluations, they still may not provide additional information, especially if the different evaluator group scores are highly correlated.[12] Ideally, these evaluations should also be completed by trained observers, but the logistics of training such a diverse group of evaluators is even more challenging than for faculty making this goal unrealistic.[13]

In an uncallibrated evaluation system, concern arises over the potential for assessment bias and likely sources of such bias. One of our particular concerns is that a halo effect may result from evaluators being unduly influenced by resident affability (ICS) and availability (PR). This concern is reflected in the well-accepted tradition of the "3 As of private practice": affability, availability, and ability, which are the factors most important to a referring physician when picking a consultant in descending order of priority. Ample evidence exists that physicians lack insight into their biases toward other physicians. It has been shown, for example, that referral patterns are not based on physician skill but on relationships.[14] A more recent study of breast cancer patients demonstrated that self-referred patients are more likely to go to high-volume, accredited centers.[15] Despite such studies, physicians still report that consultant skill and competency are the most important factors when referring patients. The other factors that influence how patients are referred may be analogous to factors that could bias the resident evaluation process. To test the hypothesis that higher scores on measures of affability (ICS) and availability (PR) would result in higher scores on the other competencies, specifically ratings of medical knowledge (MK) and patient care (PC), the following study was undertaken.

METHODS

The methods and research plan were reviewed by our institutional review board and found to be exempt. An internally developed rotational evaluation form based on the core competencies was used to evaluate residents. Each competency was divided into its essential components as defined by the ACGME, and questions were devised to assess residents' performance in each component. The faculty evaluation of resident form uses a total of 22 questions, whereas the peer evaluation forms used only 9 or 11 questions (based on PGY-level—senior or junior resident). Behavioral anchors were created for all questions to assist in homogeneity and inter-rater reliability; these anchors are displayed with the question electronically while the evaluation is being completed. Responses are recorded on a 1-point to 5-point Likert scale. Individual scores marked N/A (not applicable) or missing were recoded as no data to avoid compromising mean calculations. Formal validity and inter-rater reliability for the questions have not been established. Ongoing, informal calibration has been pursued using routine communication between the program office and evaluating faculty as well as through monthly meetings of the departmental Education Committee. All evaluations were completed online using a departmental website evaluation system. Faculty evaluators were a fixed group of senior staff physicians. Additionally, most faculty evaluations were created as a group evaluation wherein a representative faculty member from each service assimilates input from the other faculty members of the service into a single unified group evaluation. The scores for each question were combined to create an individual competency mean. The competency means were averaged and compiled. Individual question scores were also compiled. In addition, combined means of MK and PC were created, as well as for PR and ICS.

We attempted to validate our evaluation process by inviting program directors from outside our department to review the questions and behavioral prompts, which allowed for establishing face validity of our questions. We then further validated the content of the questions by having the evaluators read through the questions and determine the importance of that question in evaluating the residents, and then we used the consensus to include or exclude questions as described by Lawshe.[16]

A dataset including resident demographics and performance metrics was created. A separate dataset including a single academic year's evaluations was also constructed. To evaluate the original hypothesis, correlations were used to explore the relationships between mean competency scores. Linear regression was used to establish the predictive power of PR and ICS on MK and PC ratings. In addition to faculty evaluation scores, aggregate scores from peer evaluations and nonphysician group evaluations were also used to compare to the ICS and PR competencies (Table 1). Data were compiled in Excel 2003 (Mi-