

**Authors:**

Teresa L. Massagli, MD  
Jan D. Carline, PhD

Competency

**Affiliations:**

From the Departments of Rehabilitation Medicine (TLM), Pediatrics (TLM), and Medical Education (JDC), University of Washington, Seattle, Washington.

**Correspondence:**

All correspondence and requests for reprints should be addressed to Teresa L. Massagli, MD, Department of Rehabilitation Medicine, Mailstop W6847, 4800 Sand Point Way NE, Seattle, WA 98105.

0894-9115/07/8610-0845/0  
*American Journal of Physical Medicine & Rehabilitation*  
Copyright © 2007 by Lippincott Williams & Wilkins

DOI: 10.1097/PHM.0b013e318151ff5a

**EDUCATION & ADMINISTRATION**

## Reliability of a 360-Degree Evaluation to Assess Resident Competence

**ABSTRACT**

Massagli TL, Carline JD: Reliability of a 360-degree evaluation to assess resident competence. *Am J Phys Med Rehabil* 2007;86:845–852.

**Objective:** To determine the feasibility and psychometric qualities of a 360-degree evaluation of physical medicine and rehabilitation (PM&R) residents' competence.

**Design:** Nurses, allied health staff, and medical students completed a 12-item questionnaire after each PM&R resident rotation from January 2002 to December 2004. The items were derived from five of the six competencies defined by the Accreditation Council for Graduate Medical Education (ACGME).

**Results:** Nine hundred thirty evaluations of 56 residents were completed. The alpha reliability coefficient for the instrument was 0.89. Ratings did not vary significantly by resident gender. Senior residents had higher ratings than junior residents. A reliability of  $>0.8$  could be achieved by ratings from just five nurses or allied health staff, compared with 23 ratings from medical students. Factor analysis revealed all items clustered on one factor, accounting for 84% of the variance. In a subgroup of residents with low scores, raters were able to differentiate among skills.

**Conclusion:** Resident assessment tools should be valid, reliable, and feasible. This Web-based 360-degree evaluation tool is a feasible way to obtain reliable ratings from rehabilitation staff about resident behaviors. The assignment of higher ratings for senior residents than junior residents is evidence for the general validity of this 360-degree evaluation tool in the assessment of resident performance. Different rater groups may need distinct instruments based on the exposure of rater groups to various resident activities and behaviors.

**Key Words:** Physician Competence, 360-Degree Evaluation, Graduate Medical Education, Physical Medicine and Rehabilitation

**T**he Accreditation Council for Graduate Medical Education (ACGME) initiated the Outcome Project to increase emphasis on educational outcomes in the accreditation of residency programs.<sup>1</sup> In February 1999, the ACGME endorsed the six general competencies for residents: medical knowledge, patient care, professionalism, practice-based learning and improvement, systems-based practice, and interpersonal and communication skills. These general competencies have subcategories, resulting in a total of 28 skills for programs to teach and assess. The ACGME further collaborated with the American Board of Medical Specialties (ABMS) to produce the Assessment Toolbox, which describes 13 assessment tools that could be used by residency programs to assess physician competence.<sup>2</sup> The toolbox stratifies the assessments for each of the 28 subcompetencies at three levels: the most desirable method, the next best method, or a potential method to measure resident competence.

One of the assessments in the toolbox is a 360-degree evaluation. In the business world, 360-degree evaluations involve feedback to a person, usually a manager or executive, about how others in the workplace see him or her.<sup>3-5</sup> Typically, a survey or questionnaire is used to gather information from all coworkers (subordinates, peers, supervisors) about multiple areas of performance, such as teamwork, communication skills, decision making, or management skills. Multiple raters are used because supervisors are not able to evaluate all aspects of an employee's behaviors. It is usually a formative process to enhance behavior change, not one used to determine salary or promotion decisions.<sup>4-6</sup> Raters tend to provide more accurate and less lenient ratings when the 360-degree tool is used to give formative rather than summative feedback.<sup>4,6</sup> Once results are compiled, the ratee is expected to analyze results and develop an improvement plan. Internal consistency reliabilities as high as 0.9 have been reported in business, military, and education settings.<sup>5</sup> The 360-degree evaluation tool would seem to have some appeal to residency programs in that according to the Assessment Toolbox,<sup>2</sup> it could help to assess physician competence in all six general competencies or in 19 of the 28 skills.

Studies of internal medicine attending and resident physicians have shown that peer physicians, faculty, nurses, and patients can reliably rate physicians' humanistic behaviors.<sup>7-13</sup> It has been estimated that ratings from 10-11 peer physicians,<sup>7,9</sup> 5-15 nurses,<sup>8,12</sup> 20-50 faculty supervisors,<sup>13</sup> or 50-147 patients<sup>13,14</sup> are needed to get reliable ratings of physicians' humanistic qualities. These studies did not use 360-degree methodology; rather, they used different survey formats or rating

methods for different raters. The studies show that different groups of raters have different perspectives on physician behavior, and that ratings among physicians, nurses, and patients had variable interrater reliability.<sup>10,13</sup>

Musick and colleagues<sup>15</sup> developed a 360-degree evaluation for physical medicine and rehabilitation (PM&R) residents. Their instrument had 26 items rating overall competence, quality of patient care, personal characteristics/professionalism, and communication, derived from literature review and consultation with unit managers. It was completed at the end of each inpatient rotation by therapists, nurses, social workers, case managers, and psychologists. Four hundred twenty-one ratings were collected on 18 residents during 4 yrs. The instrument had an alpha reliability of 0.99. There were no differences in ratings by rater profession, but some differences were noted by resident gender (higher for women) and by the type of inpatient rotation (higher ratings in spinal cord injury than brain injury).

A few researchers have developed 360-degree evaluations in response to the ACGME requirement to assess specific areas of resident competence. Joshi et al.<sup>16</sup> studied a 360-degree assessment of obstetrics-gynecology residents' competency in interpersonal and communication skills. This survey was administered on one occasion to small numbers of nurses, faculty, allied health personnel, medical students, patients, and fellow residents. Some raters may have had recent exposure to the residents, and some exposures may have been more remote. Faculty and allied health personnel had high intercorrelations, but neither nurses' nor patients' ratings correlated with those of anyone else, and medical students and peer ratings were negatively correlated. The reliability of the instrument was not reported, but the results suggest that the different types of raters view the residents differently from each other. Wood and colleagues<sup>17</sup> administered a 360-degree evaluation of radiology residents' professionalism and interpersonal skills to small numbers of residents, faculty, and patients. The evaluation was completed by a patient, a resident, and a faculty member after each breast biopsy procedure, making this somewhat of a hybrid of a 360-degree evaluation and a focused or checklist observation. There was no significant correlation between resident self-ratings and patient ratings, and there was modest correlation between faculty and patient ratings, again emphasizing that people interpret interactions in different ways. The majority of the seven residents involved felt that the evaluations increased their awareness of how they interacted with patients, but the authors note that the methodology posed challenges regarding distribution and collection of forms and analysis of

results. Weigelt and colleagues<sup>18</sup> developed a 360-degree evaluation for surgery residents on intensive care unit rotations. Ten residents evaluated themselves and were also evaluated by chief residents, fellows, faculty, and nurses. The ratings were not statistically different among groups of raters (chief resident, faculty, nurses, staff, or self) for any ACGME competency. The authors felt that the instrument offered no new information about residents that was different from the standard faculty evaluation.

There are significant feasibility challenges to using 360-degree surveys in evaluation of resident performance. Collecting data using paper surveys is burdensome and expensive.<sup>13</sup> Because feedback is recall dependent, raters should be asked to evaluate residents close to the time when they have worked with the resident. This requires frequent sampling, again adding to the burden of data collection and data management. At the University of Washington, all faculty have free access to Catalyst Web Tools WebQ, a copyrighted, Web-based system that allows raters to complete surveys online. Results can be downloaded into a data file and then imported into statistical software for analysis. After studying the ACGME requirements for assessing competencies, we developed a 360-degree evaluation tool for residents that could be completed online. The purpose of this study was to assess the psychometric properties of the tool, including the internal consistency reliability and the reproducibility of the tool.

## METHODS

In October 2001, the University of Washington PM&R residency training committee (selected faculty, residents, and residency program coordinator) designed a 360-degree evaluation tool for PM&R residents. We considered who would rate the residents, then we used the toolbox list of the 19 skills to develop a 360-degree assessment of behaviors that we believed these raters could assess. We then piloted the evaluation at one institution and made final revisions to the tool before implementing the evaluation process at all training sites. A five-point Likert-type scale was used to rate the resident's skill as 5 = outstanding, 4 = very good, 3 = good, 2 = fair, and 1 = poor. An option of *not assessed* was available for each item (Appendix).

Raters included nurses, allied health professionals (social workers, psychologists, orthotists, prosthetists, vocational rehabilitation counselors, educators, physical therapists, occupational therapists, therapeutic recreation specialists, and speech/language therapists), and medical students. Raters self-selected, on the basis of their exposure to the resident, whether to complete the 360-degree tool. We excluded attending physicians, peer residents,

and patients. Attending physicians were excluded because they perform comprehensive end-of-rotation evaluations on residents, and we did not think attendings would necessarily directly observe all the types of behaviors to be rated. Peer residents were excluded because most residents had not worked on teams with other residents. Patients were excluded for feasibility reasons; for instance, prior work has estimated that 25 to as many as 147 patients are needed to achieve reliabilities of 0.7 or 0.8 on the instruments tested.<sup>13,14,19</sup> Additional challenges include language barriers, reading level of the instrument, patient access to technology for Web-based surveys and interactions between patient ratings and patient age, perceived health status, and variability in encounter setting (single outpatient *vs.* longer inpatient contact).<sup>13</sup>

Of the 19 skills identified in the toolbox as assessable by a 360-degree evaluation tool, we thought that our raters would be able to observe and rate 12 skills. These were from five competencies, excluding evaluation of medical knowledge. We constructed questions for these 12 skills (Appendix). For some items, we used the ACGME language for the competency (e.g., "demonstrates sensitivity and responsiveness to patient's culture, age, gender, and disabilities"), and for some items we modified the ACGME-suggested language to make the competency more rehabilitation specific (e.g., "work with healthcare professionals, including those from other disciplines, to provide patient-focused care" was reinterpreted as "participates in rehabilitation therapies, interventions, and patient education"). Several of the competencies as defined by ACGME are found in more than one of the six major categories. For example, the concept of working with other healthcare professionals is incorporated in *patient care*, *interpersonal and communication skills*, and *systems-based practice*. We created a single question for this, listed in the Appendix under *systems-based practice*: "works effectively as a member or leader of the team to provide patient-focused care; understands how his or her actions affect others." It could arguably be placed in either of the other two competencies. Each of the items to be rated was intended to reflect a specific type of observable skill by the resident, not a personality trait.

The 360-degree tool was piloted at one institution in December 2001. Raters were oriented to the purpose of the evaluation and asked for feedback about the ease of use of the online survey and the content of the questions. Revisions to the survey were made on the basis of this input. Clinical administrators at each of our teaching institutions were contacted to orient them to the process and to obtain e-mail list servers for the rehabilitation nurses and staff at each site. Because these e-mail

list servers are constantly modified for staff turnover, we do not know the total number of personnel contacted.

From January 2002 through December 2004, staff were contacted by e-mail at the end of each resident rotation cycle and invited to complete the online survey. The survey did not include a request for comments, because our purpose was to determine the reliability of the instrument in measuring specific behaviors, not global impressions. In the 360-degree tool, staff were asked to identify their institution and their role as either RN (nurse) or other rehab staff. They also were asked to provide their last name, with the assurance that their identity would not be disclosed to the resident. This identifier was used to examine the data for duplicate entries, in the case that staff clicked on the *submit* icon more than once. Medical students completed the same survey on paper at the end of their rotation, along with other paper-based evaluation forms. The student evaluations were anonymous. Students rated one or two residents on each rotation. Their surveys were entered into the database via the online survey by one of the researchers.

Results were downloaded into an Excel file, and one of the researchers added data for each resident, indicating the resident gender, year of training (postgraduate year [PGY] 2, 3 or 4), and whether the rotation was primarily inpatient or other (hospital consult, outpatient clinic, or electrodiagnosis studies). Statistical analyses included demographic information: resident gender, year of training, inpatient *vs.* other, and hospital site. Mean scores for the items were calculated and compared by type of rater (nurse, other staff, medical student), resident gender, year of training, type of rotation, and site. To deal with missing data, the dataset was reduced to those ratings that had at least 75% of the rating items completed. We then calculated estimated values for missing data according to an expected maximization algorithm.<sup>20</sup> This method was used over a simple mean substitution because it takes into account uncertainty about missing data related to a number of plausible solutions. Pearson product-moment correlations were calculated to assess the strength of the association between items. Principal-components analysis with varimax rotation was done to determine the factorial structure of the data. Varimax rotation is a commonly used method to increase the interpretability of factor structure if there is more than one factor found in the data.<sup>21</sup> Cronbach's alpha was calculated to determine the reliability of the scales identified by factor analysis. Generalizability decision study analyses were calculated to determine the number of individuals needed to obtain reliable ratings.<sup>22</sup> Generalizability methods provide

reliability estimates that take into account a number of sources of variability—in this case, the type of respondent as well as the individual resident. SPSS version 13.0 was used for statistical analyses. The procedures in this study were approved by the human subjects division at the University of Washington.

## RESULTS

During the three-calendar-year period, 944 ratings were submitted. The total number of possible ratings is not known, because we used mailing list servers that were independently maintained by each site. Some team members do not work with residents, and, therefore, they self-select to not respond. Visual inspection of the data revealed 14 instances in which the same rater submitted duplicate ratings on the same resident, so one of each pair was randomly deleted. Of the 930 valid ratings, 168 (18%) were from medical students, 206 (22%) were from nurses, and 556 (60%) were from other rehab staff. Medical students could have rated one or two residents, so the absolute number of students who contributed these ratings cannot be determined. The 762 ratings from nurses and other rehab staff were provided by 100 individuals.

There are a total of 28 residents in PGYs 2, 3, and 4 in the PM&R residency program at the University of Washington. This study, from January 2002 through December 2004, crossed four academic years. A total of 56 residents were evaluated, 24 women and 32 men. The distribution of ratings by training year was 466 (50.1%) in PGY 2, 278 (29.9%) in PGY 3, and 186 (20%) in PGY 4. Individual residents received between 1 and 38 ratings, with an average of 16.7 ratings per resident.

Ratings for residents on inpatient rotations accounted for 665 (71.5%) evaluations compared with 265 (28.5%) for noninpatient rotations. Residents only spend 12 of 36 mos of their training on inpatient rotations, so the larger number of inpatient ratings is likely attributable to the larger number of staff that residents work with on inpatient *vs.* other types of rotations. The number of evaluations from each hospital was not correlated with the number of residents at the site. The largest number of evaluations came from the pediatric site (331; 35.6%), likely because one of the researchers was based at that site. There were 226 (24.3%) evaluations from the university hospital, 177 (19%) from the trauma hospital, 165 (17.7%) from the Veterans Affairs hospital, and 31 (3.3%) from a private hospital.

Ratings for each of the 12 questions ranged from 1 (poor) to 5 (outstanding), and mean values ranged from 3.8 to 4.3 (Table 1). The lowest-scoring item was *works to improve the system of care*, and the highest was *demonstrates ethical behav-*

**TABLE 1** Descriptive statistics for each item

Item	Number of Responses (n)	% Missing	Mean Rating	Standard Deviation	Distribution of 958 Low Ratings (%)
Demonstrates caring and respectful behaviors with patients and families	918	1.3	4.2	1.0	6.3
Elicits information using effective questioning and listening skills	917	1.4	4.0	1.0	8.6
Effectively counsels patients, families, and/or caregivers	875	5.9	4.0	1.0	8.7
Demonstrates ethical behavior	659	29.1	4.3	0.9	2.4
Advocates for quality patient care; assists patient in dealing with system complexities	865	7.0	4.0	1.1	8.6
Sensitive to age, culture, gender, and/or disability	895	3.8	4.1	1.0	5.9
Communicates well with staff	929	0.1	3.9	1.2	12.2
Works effectively as member/leader of team; understands how own actions affect others	920	1.1	3.9	1.2	13.2
Works to improve system of care	698	25.0	3.8	1.1	9.4
Participates in rehabilitation therapies, interventions, and patient education	761	18.2	3.9	1.1	10.1
Committed to self-assessment; uses feedback for self-improvement	648	30.3	4.0	1.1	7.4
Teaches students and professionals effectively	595	36.0	4.0	1.1	7.1

Highest possible n = 930. Ratings ranged from 1 to 5. Low ratings = 1 or 2.

ior. Some skills were rated much less frequently than others. Missing data for each item ranged from 1 to 335 of 930 responses. Four items had 25% or higher missing values. The item *teaches students and professionals effectively* was rated by 99% of medical students but only by 56% of nurses and other rehab staff. This item was dropped from the remainder of the analyses because we could not compare these results across type of rater.

The dataset was further reduced to include only the 845 cases for which at least 9 of the 11 items were rated. The missing values were replaced with estimated values as described in the methods section. Item intercorrelations were quite high, ranging from 0.77 to 0.90. Factors were extracted using principal-component analysis, with varimax rotation based on mean item rating across cases. One factor obtained an eigenvalue of 9.98 and accounted for 84% of the total variance in the data. No other factor received an eigenvalue of 1 or greater. Factor loadings ranged from a low of 0.88 to a high of 0.93 for this single factor. Cronbach's alpha scale reliability for the 11-item instrument was 0.89.

Given the single factor, we further examined the data to see whether ratings were uniform for each resident across each item. For the 12 items in the 360-degree tool, the 56 residents received a total of 10,908 ratings. We defined low ratings as

1 = poor or 2 = fair. There were a total of 958 (8.8%) low ratings. We defined an outlier group as those residents who received ratings of 1 or 2 for 5–25% of their ratings. Those with fewer than 5% low ratings were viewed as random events. Those with more than 25% low ratings were viewed as poor performers. Twenty-three residents had low ratings in this outlier range. Of these, 12 residents received low ratings in relatively similar frequencies across all items. The other 11 had variability in the frequency of low ratings. For example, for one resident, 20% of ratings were low, but 50% of the low ratings were from only two items. So, 11/56 (20%) of residents had low ratings that clustered in certain items, suggesting that raters were able to differentiate among behaviors. Further evidence of the ability of raters to discriminate among items is that the frequency of low ratings varied from a low of 2.4% for *demonstrates ethical behaviors* to a high of 13.2% for *works effectively as a member/leader of the team* (Table 1).

Generalizability analyses determined that the number of ratings to achieve a reliability of 0.7 at the resident level was 3 from RNs, 2 from other rehab staff, but 13 from medical students. To achieve a reliability of 0.8, 5 ratings from RNs, 4 from other staff and 23 from medical students were needed.

The PGY 4 residents had higher ratings on all items than the more junior residents, and the dif-

**TABLE 2** Ratings by year of training and type of rotation

Item	Ratings of PGY 4 n = 172		Ratings of PGY 3 n = 242		Ratings of PGY 2 n = 431		Ratings from Noninpatient Sites n = 234		Ratings from Inpatient Sites n = 611		P Value (t test)
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Caring behaviors	4.4	0.8	4.1	1.0	4.2	1.0	4.3	0.8	4.1	1.0	0.084
Effective questioning and listening	4.2	0.8	3.9	1.1	4.0	1.1	4.1	0.9	4.0	1.1	0.081
Effective counseling	4.1	0.9	3.9	1.1	4.0	1.1	4.1	0.9	3.9	1.1	0.104
Demonstrates ethical behavior	4.2	0.8	4.1	1.0	4.1	1.0	4.2	0.8	4.1	1.0	0.343
Advocates for quality	4.2	0.9	3.9	1.1	4.0	1.1	4.1	1.0	4.0	1.1	0.146
Sensitive to age, culture, gender, and/or disability	4.2	0.8	4.0	1.0	4.1	1.0	4.2	0.9	4.1	1.0	0.294
Communicates well with staff	4.2	1.0	3.9	1.2	3.9	1.3	4.1	1.1	3.9	1.2	0.021*
Works effectively as team member/leader	4.1	1.0	3.9	1.2	3.9	1.3	4.0	1.0	3.9	1.2	0.264
Works to improve system of care	3.9	0.9	3.7	1.1	3.8	1.1	3.9	1.0	3.8	1.1	0.206
Participates in therapies and patient education	4.0	0.9	3.8	1.1	3.9	1.1	4.0	1.0	3.9	1.1	0.082
Committed to self-assessment/uses feedback	4.1	0.9	3.9	1.0	3.9	1.1	4.0	0.9	3.9	1.1	0.019*
Teaches effectively	4.2	0.9	4.0	1.1	4.0	1.2	4.2	0.9	4.0	1.1	0.008*

\* P < 0.05.

ferences were statistically significant for five items: caring and respectful behaviors, effective counseling skills, advocating for quality patient care and assisting patients in dealing with system complexities, communication with staff, and working effectively with the team as a member or leader. Ratings on outpatient/consult rotations were usually higher than on inpatient rotations and were statistically significant for three items: communication with staff, committed to self-assessment and using feedback, and teaching (Table 2). Residents primarily do inpatient rotations during PGY 2 and outpatient/consult/electrodiagnosis rotations during PGY 3 or PGY 4. The ratings at one site (trauma center) were consistently lower by 0.4–0.5 than at the other sites. Medical students were much more lenient and much less discriminating than other raters. Their average ratings were 0.5–0.8 points higher, and their standard deviations were lower than for other staff. They tended to rate the residents as excellent or very good in most areas. Ratings did not vary significantly by gender of the resident.

## DISCUSSION

Ideally, assessment instruments should be reliable, valid, feasible, and provide valuable information about whatever is being measured. Reliability of the assessment process is enhanced by using multiple observations over time and by using multiple observers. Although a 360-degree evaluation may not be able to evaluate all aspects of competence, it does use multiple observers and can be done repeatedly to assess change in skills. The alpha scale reliability of our instrument was quite high at 0.89, indicating a high degree of internal consistency. We believe that the increase in rating scores with more advanced levels of training (year of training or outpatient *vs.* inpatient rotation) is evidence for the general validity of the tool in assessing resident competence.

We designed our 12-item, 360-degree evaluation instrument to assess five of the competencies (all except medical knowledge) endorsed by the ACGME. Although it is believed that physician competence is multidimensional and that no single tool will be able to assess all aspects of competence, ratings from multiple observers in this study resulted in ratings of a single unified factor, even though the instrument had been carefully designed to address several competencies. We were initially disappointed to see that factor analysis identified only one factor. It could be that we did not ask enough or the right questions to define each area of competence. It could also be that our raters were not adequately trained to distinguish among the items and that they tended to rate the residents according to a single gestalt sense of their skills. However, our outlier analysis, looking at low rat-

ings of poor or fair, showed that raters did not assign these low ratings uniformly across items, and we feel that such outlier low ratings may indicate specific behaviors of concern.

As stated in the Methods section, the 28 skills included by ACGME in the six competencies show considerable overlap, and no research has been done to demonstrate that they are indeed distinct factors. A review of the literature on ratings of physician performance, whether by peers, nurses, patients, or program directors, reveals that generally two factors account for 78–90% of the variance in ratings.<sup>7,9,23,24</sup> These two dimensions are knowledge/clinical judgment and interpersonal skills. We specifically designed our 360-degree tool to exclude questions related to medical knowledge or clinical judgment, so in retrospect it is not so surprising that we were able to identify only one factor. But, the dimensions of cognitive skills and interpersonal skills are composed of many different behaviors in many different situations. A physician could be substandard in some areas, proficient in others, and viewed overall as competent. Our analysis shows that raters were able to discriminate among specific behaviors and that, when interpreting scores for residents, both the individual items and the overall performance should be considered.

The administration of the 360-degree evaluation to staff and students was feasible, using a Web-based survey and data-management tool. We only needed evaluations from as few as four or five nurses/allied health staff to have a resident evaluation with a reliability coefficient of 0.8. Unfortunately, medical students were less discriminating, and the number of student evaluations needed for a reliable rating of a resident is impossibly high for our program. We chose not to include patients in this assessment for the feasibility issues previously described.

This study only addressed the psychometric properties of our tool. The next step will be to incorporate it into resident evaluation as a 360-degree evaluation process. The tool itself represents only multirater feedback. At least one residency program felt that the multirater feedback data provided by the 360-degree evaluation format did not add any new information that the program director did not already know from attending evaluations.<sup>18</sup> However, a benefit of the process is to help the resident understand how other members of the team view his or her skills. An effective 360-degree evaluation process must also include a formative process whereby the residents can review the data received and develop a plan for improvement if needed. To achieve a change in behavior, the data must be of high quality, and the resident must be willing to accept it as formative feedback from colleagues who have his or her best interests

at heart. We have found that our data are reliable and valid. The next step is to get the resident to use it. Reflection, development of an action plan, and reassessment at a later date make the resident a more active participant in the evaluation process and may indeed contribute something new and valuable to resident education and evaluation. Peer and patient input and self-evaluation are currently required by some boards for maintenance of certification, so beginning this process in residency may also help prepare physicians for these types of evaluation procedures.

As we conducted the study, we did note some differences using a 360-degree evaluation in residency training as compared with in a business environment. Residents are not managers working in a stable environment. They change rotations frequently, and their roles change depending on assignments. It may be difficult for a resident to identify behaviors to change when his or her role changes, and it may be difficult to measure change because there are new raters at the new sites. It may be difficult to include peers if residents are on individual assignments. A single set of questions for all raters who interact with residents may not be sufficient. We found that medical students were more likely to rate the teaching skills of residents than any other rater. Nurses and other rehab staff likely have fewer opportunities to observe the residents' teaching skills. Forms may need to be developed that recognize the availability of behaviors or incidents for raters to observe. In the corporate world, the 360-degree tool is most often used for formative purposes, not summative ones, but in residency, many evaluation tools have some summative focus. The data from a 360-degree tool may be difficult to use for important summative decisions such as nonreappointment, because the feedback to the resident is anonymous. However, the resident's success at developing an action plan and improving behavior could be judged as part of summative decision making.

## CONCLUSION

We found that our 360-degree evaluation tool is reliable, valid, and feasible to administer to hospital-based staff. The alpha reliability coefficient was 0.89. Senior residents received higher ratings than junior residents, as would be expected with progressive skill development during training. The Web-based format was efficient and easy to use. The 360-degree evaluation seems ideally suited to the field of PM&R, given the team-oriented nature of our discipline and the fact that residents work side by side with many other health professionals. The 360-degree evaluation tool did not measure five distinct dimensions of competence, but it is one tool in the evaluation toolbox that can be used to assess per-

formance, from which we can infer competence. A single 360-degree evaluation form may not work for all types of raters if they do not share similar opportunities to observe resident skills.

## APPENDIX

The following competencies were evaluated by raters on a five-point scale in which 5 = outstanding, 4 = very good, 3 = good, 2 = fair, and 1 = poor; or, they were designated as *not assessed*.

### Patient Care

Demonstrates caring and respectful behaviors (verbal and nonverbal) with patients and families  
Effectively counsels patients, families, and caregivers

Participates in rehabilitation therapies, interventions, and patient education

### Interpersonal and Communication Skills

Elicits information using effective questioning and listening skills

Communicates well with staff

### Professionalism

Demonstrates ethical behavior pertaining to provision or withholding of care, informed consent, and confidentiality

Demonstrates sensitivity and responsiveness to patient age, culture, gender, and disability

### Practice-Based Learning and Improvement

Committed to self-assessment; uses feedback for self-improvement

Teaches students and professionals effectively

### Systems-Based Practice

Advocates for quality patient care; assists patients in dealing with system complexities

Works to improve the system of care

Works effectively as a member or leader of the team to provide patient-focused care; understands how his or her actions affect others

## REFERENCES

1. Accreditation Council for Graduate Medical Education: Outcome Project. Available at: [www.acgme.org/outcome](http://www.acgme.org/outcome). Accessed July 26, 2007.
2. Accreditation Council for Graduate Medical Education/American Board of Medical Specialties: Toolbox of Assessment Methods, version 1.1. Available at: [www.acgme.org/outcome/assess/toolbox.pdf](http://www.acgme.org/outcome/assess/toolbox.pdf). Accessed July 26, 2007.
3. Info-line: How to build and use a 360 degree feedback system. *Am Soc Train Dev* 1998;9508:1-13
4. Brett JF, Atwater LE: 360 degree feedback: accuracy, reactions and perceptions of usefulness. *J Appl Psychol* 2001; 86:930-42
5. Collins ML: *The Thin Book of 360 Degree Feedback: A Manager's Guide*. Plano, TX, Thin Book Publishing Company, 2000
6. Bracken DW, Dalton MA, Jako RA, McCauley CD, Pollman VA: *Should 360 Degree Feedback Be Used Only for Developmental Purposes?* Greensboro, NC, Center for Creative Leadership Publications, 1997
7. Ramsey PG, Wenrich MD, Carline JD, et al: Use of peer ratings to evaluate physician performance. *JAMA* 1993;269: 1655-60
8. Wenrich MD, Carline JD, Giles LM, Ramsey PG: Ratings of the performances of practicing internists by hospital-based registered nurses. *Acad Med* 1993;68:680-7
9. Ramsey PG, Carline JD, Blank LL, Wenrich MD: Feasibility of hospital-based use of peer ratings to evaluate the performance of practicing physicians. *Acad Med* 1996;71:365-70
10. McLeod PJ, Tamblyn R, Benaroya S, Snell L: Faculty ratings of resident humanism predict patient satisfaction in ambulatory clinics. *J Gen Intern Med* 1994;9:321-6
11. Kaplan CB, Centor RM: The use of nurses to evaluate house officers' humanistic behavior. *J Gen Intern Med* 1990;5: 410-4
12. Butterfield PS, Mazzaferri EL: A new rating form for use by nurses in assessing residents' humanistic behavior. *J Gen Intern Med* 1991;6:155-61
13. Wooliscroft JO, Howell JD, Patel BP, Swanson DB: Resident-patient interactions: the humanistic qualities of internal medicine residents assessed by patients, attending physicians, program supervisors, and nurses. *Acad Med* 1994; 69:216-24
14. Nelson EC, Gentry MA, Mook KH, Spritzer KL, Higgins JH, Hays RD: How many patients are needed to provide reliable evaluations of individual clinicians? *Med Care* 2004;42: 259-66
15. Musick DW, McDowell SM, Clark N, Salcido R: Pilot study of a 360-degree assessment instrument for physical medicine and rehabilitation residency programs. *Am J Phys Med Rehabil* 2003;82:394-402
16. Joshi R, Ling FW, Jaeger J: Assessment of a 360-degree instrument to evaluate residents' competency in interpersonal and communication skills. *Acad Med* 2004;79:458-63
17. Wood J, Collins J, Burnside ES, et al: Patient, faculty, and self-assessment of radiology resident performance: a 360 degree method of measuring professionalism and interpersonal/communication skills. *Acad Radiol* 2004;11:931-9
18. Weigelt JA, Brasel KJ, Bragg D, Simpson D: The 360-degree evaluation: increased work with little return? *Curr Surg* 2004;61:616-26
19. Swanson DB, Webster GD, Norcini JJ: Precision of patient ratings of residents' humanistic qualities: how many items and patients are enough? in Bender W, Hiemstra R, Scherpbler AJA, Zwiestra RP (Eds): *Teaching and Assessing Clinical Competence*. Groningen, the Netherlands, Boek-Werk Publications, 1990
20. Schafer JL: NORM: Multiple Imputations of Incomplete Multivariate Data Under a Normal Model, version 2.03 [software]. Available at: <http://www.stat.psu.edu/~jls/misoftwa.html>. Accessed July 26, 2007
21. Nunnally JC: *Psychometric Theory*. New York, NY, McGraw Hill Book Company, 1978
22. Shavelson RJ, Webb NM: *Generalizability Theory*. Newbury Park, CA, Sage Publications, 1991
23. 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: 549-56
24. Paolo A, Bonaminio G: Measuring outcomes of undergraduate medical education: residency directors' ratings of first year residents. *Acad Med* 2003;78:90-5