

Use of a 360-Degree Evaluation in the Outpatient Setting: The Usefulness of Nurse, Faculty, Patient/Family, and Resident Self-Evaluation

NICOLE CHANDLER, MD
GAVIN HENDERSON, MD, PhD
BRITTANY PARK, BS
JULIE BYERLEY, MD, MPH
WALLACE D. BROWN, MD
MICHAEL J. STEINER, MD

Abstract

Background Faculty have traditionally evaluated resident physician professionalism and interpersonal skills without input from patients, family members, nurses, or the residents themselves. The objective of our study was to use “360-degree evaluations,” as suggested by the Accreditation Council for Graduate Medical Education (ACGME), to determine if nonfaculty ratings of resident professionalism and interpersonal skills differ from faculty ratings.

Methods Pediatrics residents were enrolled in a hospital-based resident continuity clinic during a 5-week period. Patient/families (P/Fs), faculty (MD [doctor of medicine]), nurses (RNs [registered nurses]), and residents themselves (self) completed evaluator-specific evaluations after each clinic session by using a validated 10-item questionnaire with a 5-point Likert scale. The average Likert score was tallied for each questionnaire. Mean Likert scale scores for each type of rater were compared by using analysis of variance, text with pair-wise comparisons when appropriate. Agreement between rater types was measured by using the Pearson correlation.

Results A total of 823 evaluations were completed for 66 residents (total eligible residents, 69; 95% participation). All evaluators scored residents highly (mean Likert score range, 4.4 to 4.9). However, MDs and RNs scored residents higher than did P/Fs (mean scores: MD, 4.77, SD [standard deviation], 0.32; RN, 4.85, SD, 0.30; P/F, 4.53, SD, 0.96; $P < .0001$). MD and RN scores also were higher than residents’ self-evaluation scores, but there was no difference between self-scores and P/F scores (average resident self-score, 4.44, SD, 0.43; $P < .0001$ compared to MD and RN; $P = .19$ compared to P/F). Correlation coefficients between all combinations of raters ranged from -0.21 to 0.21 and none were statistically significant.

Conclusion Our study found high ratings for resident professionalism and interpersonal skills. However, different members of the health care team rated residents differently, and ratings are not correlated. Our results provide evidence for the potential value of 360-degree evaluations.

Nicole Chandler, MD, is a Hospitalist, Division of General Pediatrics, Department of Pediatrics at The Children’s Hospital of Philadelphia; **Gavin Henderson, MD, PhD**, is a Fellow, Division of Neonatology, Department of Pediatrics at the University of California, San Francisco; **Brittany Park, BS**, is a Physician Assistant Student, Department of Physician Assistant Studies, College of Allied Health Sciences, East Carolina University; **Julie Byerley, MD, MPH**, is an Associate Professor, Division of General Pediatrics and Adolescent Medicine, Department of Pediatrics, The University of North Carolina at Chapel Hill School of Medicine; **Wallace D. Brown, MD**, is Professor, Division of General Pediatrics and Adolescent Medicine, Department of Pediatrics, The University of North Carolina at Chapel Hill School of Medicine; and **Michael J. Steiner, MD**, is Assistant Professor, Division of General Pediatrics and Adolescent Medicine, Department of Pediatrics, The University of North Carolina at Chapel Hill School of Medicine.

We would like to thank the patients and families, nurses, faculty, and ancillary staff in the University of North Carolina Pediatric Outpatient Clinic for their participation in this study; the Odum Institute at the University of North Carolina for biostatistical consultation and support; and Dr Kenneth Roberts, Dr Harvey Hamrick, and Dr Eliana Perrin for their review of the article before submission; and Dr Raksha Joshi (MD, MS, MBBA, MBA is assistant professor of Obstetrics and Gynecology at Drexel University College of Medicine) for use of the evaluation tool.

The authors declare that they have no competing or conflicting interests.

Corresponding author: Nicole Chandler, MD, The Children’s Hospital of Philadelphia, 34th and Civic Center Boulevard, Philadelphia, PA 19104, 919.724.3624, chandlern@email.chop.edu

Received January 17, 2010; revisions received March 20, 2010, and May 18, 2010; accepted June 6, 2010.

DOI: 10.4300/JGME-D-10-00013.1

Editor’s Note: The online version of this article contains the 4 questionnaires (nurse, physician faculty, self, and patients) used to evaluate residents in this study.

Background

Professionalism and interpersonal and communication skills are important to allow physicians to gain and maintain the trust of patients and work as a team with other health care professionals. These skills contribute to patient satisfaction with their encounter; moreover, residents with low skills in these areas are more likely to have future disciplinary actions taken against them by medical boards.¹ The importance of teaching these skills in medical education has become an area of increasing interest among medical educators.^{1,2}

The ACGME Outcome Project lists interpersonal and communication skills among the 6 core competencies and states that “residents must be able to demonstrate interpersonal and communication skills that result in effective information exchange and teaming with patients,

their patients' families, and professional associates."² Accurate assessments of these skills during medical training could potentially allow for remediation of identified deficiencies.

Traditionally, evaluation of resident professionalism and interpersonal and communication skills has been completed by faculty evaluators. These evaluations are limited for 2 reasons: often, faculty have not directly observed residents as they interact with patients and families,^{3,4} and faculty evaluations do not consider other perspectives from nurses, patients, or residents' self-evaluation of their skills. The ACGME has suggested the use of multisource feedback and 360-degree evaluation systems as 1 method, using multiple evaluators for assessing resident skills in a variety of areas including professionalism and interpersonal and communication skills.⁵ This evaluation system consists of "measurement tools completed by multiple people in a person's sphere of influence."⁵ Previous studies^{6–8} have shown that this system can enhance the evaluation of interpersonal and communication skills in residents.

Feedback given to a resident by evaluators after interactions not only provides the resident with information to improve skills but also may provide insight into deficiencies that the resident does not realize exist.⁹ Another study⁶ analyzed the use of 360-degree evaluations for pediatrics residents on the inpatient service at a children's hospital, finding that ratings by nurses were different from those provided by other raters. A corollary finding was the absence of significant difference between ratings of resident skills provided from faculty physicians and patient families. This finding raises potentially important questions about the importance of including families in resident evaluations and whether faculty feedback can serve as a surrogate for ratings directly from families.⁶

Whereas most pediatric medical care and a significant amount of residency training occur in outpatient settings,¹⁰ the study by Brinkman et al⁶ involved only evaluation of inpatient care. The assessment of resident professionalism, communication, and interpersonal skills in these 2 settings may be different. First, the outpatient continuity clinic setting is the primary place where residents provide longitudinal care of children over time and thereby develop a long-term relationship with individual patients and their families. Secondly, a large percentage of the outpatient care of children centers on well-child care. Physician skills and parental expectations for well-child care may be very different from those for ill care, particularly when children are ill enough to necessitate hospitalization. For all these reasons, an assessment of 360-degree evaluations in outpatient care is important. Different members of the health care team may have unique perspectives of physicians' skills, and these differing perspectives may be useful feedback for practicing physicians.¹¹

The purpose of our study was to compare the evaluations of pediatrics residents in the outpatient setting

from 360-degree perspectives, including evaluations from faculty, nursing staff, and patient/families and resident physician self-evaluations. Since using multiple evaluators to assess resident skills is a program requirement, we aimed to design our study in a realistic setting and time period that would be easily replicable annually or biannually for resident assessment. We hypothesized that in this outpatient setting, patient families would rate resident professionalism and communication differently than faculty and nursing staff. Specifically, given the findings in the study by Brinkman et al, we hypothesized that nurses (RNs [registered nurses]) would rate residents lower than would faculty (MDs [doctors of medicine]), patient/families (P/Fs), and self.

Methods

Subjects

This cross-sectional study enrolled resident physicians, after verbal consent during sessions of their continuity clinic, and took place during consecutive days over a 5-week period in June 2008. All eligible residents agreed to participate. Residents who did not have clinic during the time period were not eligible. Each subject was evaluated by attending physicians (MD), nursing staff (RN), patient families (P/F), and by resident self-evaluation (self) for resident physician professionalism during outpatient continuity clinic appointment encounters.

Setting

The study took place at the University of North Carolina Child and Adolescent General Clinic. Pediatrics residents (first–third year) and combined medicine-pediatrics program residents (first year) from the University of North Carolina pediatrics program were included in the study. Approximately 6000 children receive their primary care in this clinic, with approximately 10 000 visits per year. The clinic is staffed by 9 supervising physicians and a consistent nursing staff of 2 registered nurses, 1 licensed practical nurse, and 4 nursing assistants.

Each patient was initially seen by a resident physician. After obtaining a history and performing an examination, the resident or intern discussed the case with the supervising physician. The level of involvement of the supervising physician varied depending on the complexity of the case and the experience level of the resident.

Measurement Tool

The evaluation used in this study was a modified version of a tool designed by Joshi et al⁸ in a study assessing 360-degree evaluations in obstetrics and gynecology residents. The tool is a 10-item questionnaire designed to evaluate the interpersonal and communication skills of the residents. The tool was modified for each evaluator type in the study, with a total of 4 distinct questionnaires (MD, RN, P/F, and self) (online supplement). Each question had a

5-point Likert scale score with a highest possible score of 5. Questions were graded on frequency of occurrence of an observed behavior. The lowest score (0) was described as “never” and the highest score (5) was described as “always.” Evaluators were instructed to complete each question on the 10-item evaluation survey. The survey did not ask for personal identifiers of the evaluators. The surveys were labeled by residents’ last name and category of evaluator.

Evaluator Recruitment

The nursing and attending evaluation packets were distributed at the beginning of each half-day session of clinic by a study investigator. The evaluation packets included an evaluation form for each resident in clinic on the study day. The RN and MD evaluators were asked to evaluate each resident they worked with during each half-day clinic. The P/F evaluations were given to English-speaking patient/families by an RN when they were being checked in for their appointment. The P/Fs filled in the name of the resident physician seeing them in clinic on the study day, but did not identify themselves. At the end of the visit, the P/F ratings of the resident physicians were either collected by the nursing staff, or they were anonymously deposited in a collection box. The length of time that each evaluator had to get to know a specific resident varied. In general, the RN and MD would have known the residents throughout the residency training and the P/F may have known the residents for a period ranging from 1 clinic visit to the entire residency. The residents also completed self evaluations for each clinic half-day. The completed self-evaluations were placed in an envelope that was collected by a study investigator at the end of each study day.

Data Analysis

The surveys were collected and the evaluator type was identified (P/F, MD, RN, and self) and entered into a computerized database (Microsoft Excel 2003, Redmond, WA). The score for each of the 10 items was summed, and a mean item score was calculated for each tool. Combining all 10 items, each with a score from 0 to 5, created a variable that could be accurately analyzed by using means and nonparametric analysis. If individual items from the 10-item scale were omitted by an evaluator, a mean score was calculated from the completed items. A mean score was then calculated across each type of evaluator and each level of training. Mean Likert scale scores were compared by using mixed analysis of variance with random subject effects and pair-wise comparisons when appropriate. Intraclass correlation was calculated between evaluators to measure agreement. All statistical analysis was done by using SAS (Statistical Analysis Software, Cary, NC). The full study protocol was submitted to the Health Science Institutional Review Board for the University of North Carolina and a waiver of approval was granted.

Results

Sixty-six pediatrics residents, 95% of the 69 eligible residents, were enrolled and evaluated on 22 of 25 consecutive days. A total of 836 evaluations were completed by P/Fs ($n = 83$), MDs ($n = 173$), RNs ($n = 489$), and self ($n = 78$). Of these, 21 evaluations had to be excluded from the study owing to insufficient data (2.5%), with 18 of these forms from RNs and 3 from MDs.

All evaluators scored residents highly (mean score ranged from 4.4 to 4.9). However, MDs and RNs scored subjects higher than did P/Fs (mean scores: MD, 4.77, SD [standard deviation], 0.32; RN, 4.85, SD, 0.30; P/F, 4.53, SD, 0.96; $P < .0001$). The resident self-evaluation scores were also lower than the MD and RN evaluation scores (self, 4.44, SD, 0.43; $P < .0001$). The self scores and P/F scores did not differ statistically ($P = .19$).

The evaluator responses were not correlated (intraclass correlation of 0.16; 95% confidence interval, 0–0.40). Residents were scored highly, regardless of postgraduate level (PL) year (PL-1, 4.75, SD, 0.43; PL-2, 4.76, SD, 0.47; PL-3, 4.77, SD, 0.45; $P = .65$).

Discussion

The ACGME requires evaluation of professionalism and interpersonal and communication skills, which are 2 of the 6 core competencies.² In this study, we compared assessments by 4 different raters that would be part of 360-degree evaluations of pediatrics residents in continuity clinic. In this study, all residents scored highly, regardless of category of evaluator (RN, MD, P/F, and self). There are many potential explanations for these overall high scores. First, it is possible that P/F raters are highly satisfied with their care in this setting and that resident physicians actually communicate well and display high professionalism. A previous continuity clinic research network study confirmed that patients are highly satisfied with their care in this setting.¹² Another possibility is that residents in the same residency program have similar skills and, when compared to each other, all are rated highly. It would be difficult for residents or attending physicians to rate residents by some type of general physician “gold standard” instead of rating residents on the basis of comparisons within the same program. A study by Calhoun et al,¹³ performed more than 20 years ago, used standardized patients and trained evaluators for assessing internal medicine residents’ performance. In Calhoun’s study, two-thirds of standardized patient evaluators correlated moderately to strongly with the trained evaluators.¹³ However, since we sought to replicate real-world application of 360-degree evaluations in a working clinic setting, we did not pursue standardized patients.

Although all residents scored highly, our study found that P/Fs evaluated residents lower than did MDs and RNs. This difference is particularly important because physicians

need to demonstrate the attributes of professionalism and good communication toward families. Improved skills both toward P/Fs and toward RNs will potentially improve care. The difference in scores given by patients and families may reflect a difference in the way that residents interact with patients compared to members of the health care team. However, it may also be the result of different standards that patients have of physicians compared to evaluators who work in the health care field (RN, MD, and self). It should be noted that although statistically significant, this degree of difference on a Likert scale may not represent a clinically important difference.

Our study findings differ from Brinkman et al.⁶ Our results indicate that P/Fs scored residents lower than MDs and RNs. In the Brinkman et al faculty and patients gave similar ratings, and nurses rated residents lower than other raters. Davis⁷ found similar differences between raters of obstetrics and gynecology residents, finding that faculty and peer MD evaluations were correlated, but nurses' ratings were not correlated to faculty ratings. The nurses in this study also showed low intraclass correlation within their group; the nurse group included nurses from both inpatient and outpatient areas. The source of this within-group variation among nurses may be due to the different settings. The same difference may explain why our findings differed from those of Brinkman et al⁶ in the inpatient setting. It is likely that nurses and physicians who get to know each other over the 3 longitudinal years of continuity clinic develop a different relationship from that of nurses and physicians during a brief inpatient rotation.

The purpose of the 360-degree evaluations is to enhance the evaluation process and provide additional information for direct feedback and teaching. If all evaluators rated residents similarly, there would be no additional information gained from 360-degree evaluations, as opposed to the traditional individual faculty evaluations. However, in our study, different evaluators rated residents differently and in ways that were not correlated. Obtaining other perspectives beyond faculty evaluations may provide a more complete view of resident performance in different settings.

It was initially hypothesized that postgraduate-level year would affect interpersonal and communication skills and that residents would improve in these skills in each year of training. One reason that third-year residents may not have scored more highly than interns is that this is the first year that multisource feedback has been used in the residency. Therefore, the third-year residents probably were never given feedback from patients' families and nurses and therefore may not have made the expected improvements during residency. Brinkman et al¹⁴ confirmed this possibility when they demonstrated that a feedback intervention after multisource evaluation (ie, 360-degree method) improved communication skills and professional behavior of pediatrics residents. In contrast, Weigelt et al¹⁵ concluded that the 360-degree evaluation system was more time

intensive to conduct and did not enhance the evaluation of surgical residents. These disparate results suggest the need for further investigation.

A significant limitation in our study is the evaluation tool used. The tool included questions that evaluated behavior that may not have occurred during the resident-evaluator interaction. Specifically, question 10, "Resident apologizes to you for inappropriate behavior on his/her part," on the RN and MD evaluations and the resident self-evaluations may not be applicable to most of the interactions. This particular question was most likely to be answered by evaluators as "N/A" rather than given a score. However, some evaluators did respond to this question with a numbered answer. It is not clear if these cases involved the resident apologizing for an inappropriate behavior or if the evaluator provided a response regardless of the interaction. Additionally, some questions were adjusted for different evaluators, which might have limited the ability to compare scores across different evaluator types. For example, question 4 asked nurses to respond to the statement, "Resident is courteous and polite when called/answering beep"; faculty were asked, "Resident is courteous and polite when interacting with you"; patients were asked, "Resident is courteous, polite, and sensitive to my feelings and respects my desires." Different questions may be interpreted differently by different types of evaluators. Removing these questions from the final results did not change the lack of agreement between different raters (data not shown). Another limitation of the study was the number of evaluators per resident included in the study. Wenrich et al¹¹ suggested that approximately 10 to 15 ratings from nurses would provide useful assessment and that a generalizability coefficient of 0.8 would require 11.4 to 23.9 nurse respondents. In contrast, our clinic has a maximum of 7 potential nurse evaluators and 9 potential attending evaluators. Although increasing the number of evaluators may have improved the reliability of the study, it would not have reflected the practical application of 360-degree evaluations in a resident clinic. It was also not possible to determine the percentage of eligible raters that completed the surveys because we tried to maintain anonymity of evaluators completing surveys.

Non-English-speaking families were excluded from our study. This was done to avoid the potentially interacting factor of the interpreter on the communication between the resident and the family, but it is an important limitation that may affect external validity. The impact of interpreter services during patient interactions is a topic that merits further evaluation. Another potential limitation is that our study took place during a concentrated period of time. Ideally, feedback about these interpersonal skills should use information collected over a longer period of time. Lastly, despite our attempts to use an anonymous collection box for the family survey, families may have been fearful that the results would not remain anonymous.

Conclusions

Our study of 360-degree evaluations of pediatrics residents in the outpatient setting suggests that different stakeholders in the health care interaction have different perspectives on the professionalism and interpersonal and communication skills of the resident physicians. The study reflected a practical use of multiple evaluator feedback in a residency program that can be replicated annually or biannually. The results suggest that 360-degree evaluations that incorporate multiple perspectives on care might provide additional useful information. The information obtained from 360-degree evaluations can guide feedback to residents on their interpersonal and communication skills and may lead to improved patient care.

References

- 1 Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. *N Engl J Med*. 2005;353(25):2673–2682.
- 2 Accreditation Council for Graduate Medical Education. Advancing education in interpersonal and communication skills: an educational resource from the ACGME Outcome Project. Available at: <http://www.acgme.org/outcome/assess/toolbox.asp>. Accessed February 22, 2009.
- 3 Carraccio C, Englander R, Wolfsthal S, Martin C, Ferentz K. Educating the pediatrician of the 21st century: defining and implementing a competency-based system. *Pediatrics*. 2004;113(2):252–258.
- 4 Howley LD, Wilson WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. *Acad Med*. 2004;79(3):276–280.
- 5 Accreditation Council for Graduate Medical Education. 360-Degree evaluation instrument: an educational resource from the ACGME Outcome Project. In *Toolbox of Assessment Methods*. Version 1.1. September 2000;3. Available at: <http://www.acgme.org/Outcome/assess/Toolbox.pdf>. Accessed February 22, 2009.
- 6 Brinkman WB, Geraghty SR, Lanphear BP, et al. Evaluation of resident communication skills and professionalism: a matter of perspective? *Pediatrics*. 2006;118(4):1371–1379.
- 7 Davis JD. Comparison of faculty, peer, self, and nurse assessment of obstetrics and gynecology residents. *Obstet Gynecol*. 2002;99(4):647–651.
- 8 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(5):458–463.
- 9 Woolliscroft JO, TenHaken J, Smith J, Calhoun JG. Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Acad Med*. 1993;68(4):285–294.
- 10 McBurney PG, Moran CM, Ector WL, Quattlebaum TG, Darden PM. Time in continuity clinic as a predictor of continuity of care for pediatric residents. *Pediatrics*. 2004;114(4):1023–1027.
- 11 Wenrich MD, Carline JD, Giles LM, Ramsey PG. Ratings of the performances of practicing internists by hospital-based registered nurses. *Acad Med*. 1993;68(9):680–687.
- 12 Krugman SD, Racine A, Dabrow S, et al. Measuring primary care of children in pediatric resident continuity practices: a Continuity Research Network study. *Pediatrics*. 2007;120(2):e262–e271.
- 13 Calhoun JG, Woolliscroft JO, ten Haken JD. Internal medicine house officers' performance as assessed by experts and standardized patients. *J Med Educ*. 1987;62(9):754–760.
- 14 Brinkman WB, Geraghty SR, Lanphear BP, et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med*. 2007;161(1):44–49.
- 15 Weigelt JA, Brasel KJ, Bragg D, Simpson D. The 360-degree evaluation: increased work with little return? [discussion in *Curr Surg*. 2004;61(6):627–628] *Curr Surg*. 2004;61(6):616–626.