

## Using the Communication Assessment Tool in Family Medicine Residency Programs

Linda Myerholtz, PhD; Lynn Simons, PsyD; Sumi Felix, MD; Tuan Nguyen, MD; Julie Brennan, PhD; Ana Rivera-Tovar, PhD; Pat Martin, MA, LPCC; Jeri Hepworth, PhD; Gregory Makoul, PhD

**Background and Objectives:** *The Communication Assessment Tool (CAT), developed by Makoul et al assesses patient perceptions of physicians' interpersonal and communication skills. The objective of this study was to gather initial benchmarking data for the use of the CAT in family medicine residency programs. Methods:* Data were collected from patients seeing 127 residents from six family medicine residency programs. A total of 1,931 patients completed the paper and pencil version of the CAT following an appointment with a resident; 1,880 of the CAT forms met the inclusion criteria for analysis. **Results:** The overall mean percentage of items from which residents were rated as excellent was 69.7%. Significant differences were found in the overall percentage of items rated as excellent based on training year, with PGY-1 (77.0%) residents being rated significantly higher than PGY-2 (69.5%) and PGY-3 (68.1%) residents. There were no significant differences found in the overall percentage of items rated as excellent based on the native language or gender of the residents. **Conclusions:** This initial benchmarking data can allow family medicine residency programs to compare the performance of their residents with the performance of residents from other programs. We recommend that the results of the CAT be used as both an evaluative and learning tool.

(Fam Med 2010;42(8):567-73.)

The Communication Assessment Tool (CAT), developed by Makoul et al, assesses patient perceptions of physicians' interpersonal and communication skills.<sup>1</sup>

The Accreditation Council for Graduate Medical Education (ACGME) has identified interpersonal and communication skills as a core competency for all physicians,<sup>2</sup> and residency programs have the challenge of evaluating these skills. Family medicine programs have used a variety of strategies for skills assessment, including global rating forms, behavioral checklists for direct or video observations, objective structured clinical

examinations (OSCEs), 360-degree evaluations, and patient satisfaction surveys.<sup>3-7</sup> No one evaluation method will truly be able to capture a physician's competence in the complex behaviors that comprise interpersonal and communication skills, and multimodal evaluations are needed.<sup>3,8,9</sup> The patient's experience is of vital importance when evaluating physician interpersonal and communication skills<sup>3</sup> and should be included in any comprehensive evaluation system.

While there are many surveys of patient interactions with physicians, the psychometric rigor of their development is often unknown.<sup>1,10</sup> In addition, many patient surveys mix treatment satisfaction items with communication skill items, combine multiple communication skill elements into single items, and/or ask patients to rate their satisfaction over a period of time.<sup>1</sup> Such concerns often make it difficult to use these tools to provide concrete feedback on basic communication skills to physicians in a manner that could facilitate physician learning and possible behavior change.

The Communication Assessment Tool (CAT) was developed to assess patients' perceptions of the in-

---

From the Mercy Health Partners Family Medicine Residency Program, Toledo, OH (Drs Myerholtz, Felix, and Nguyen); Midland Family Medicine Residency, Midland, MI (Dr Simons); St. Luke's Hospital/University of Toledo Family Medicine Residency (Dr Brennan); York Hospital Family Practice Residency Program, York, PA (Dr Rivera-Tovar); Grant Family Medicine Residency, Columbus, OH (Ms Martin); University of Connecticut Family Medicine Residency Program (Dr Hepworth); and Saint Francis Hospital and Medical Center, Hartford, CT (Dr Makoul).

terpersonal and communication skills of physicians while addressing the previously noted limitations of other patient satisfaction tools.<sup>1</sup> The team that developed the CAT documented all steps in the process to create a reliable and valid instrument that can be easily implemented in physicians' practices and training environments. Patients are asked to reflect only on the encounter that they just had with the physician, and the CAT items focus on basic communication skill elements.

The objective of this study was to gather initial benchmarking data for the use of the CAT in family medicine residency programs. In addition, we also examined differences in patient ratings of family medicine residents in six programs, based on the resident's year in training, gender of the resident, and the native language of the resident (native English speaking versus non-native English speaking).

## Methods

A convenience sample of six family medicine residency programs agreed to participate in this study. Institutional Review Board approval was obtained at each of the participating sites. Data were collected during November and December of 2008.

### *Selection and Description of Participants*

Patients who attended an appointment with a family medicine resident at one of the participating outpatient family medicine residency clinical sites were asked to complete a paper-and-pencil version of the CAT. The patients were given the CAT at the end of their appointment by support staff. Support staff informed the patient that completion of the survey was voluntary and that their responses were anonymous and confidential. A Spanish version of the CAT was offered to patients who preferred to complete the survey in this language. The support staff instructed the patient to return the completed CAT to a secure location in the office waiting room, and resident physicians did not have access to surveys at the time of completion.

### *Resident and Program Characteristics*

All participating residency programs were located in the Midwest or East Coast (Ohio, Michigan, Pennsylvania, and Connecticut). Three of the residency programs were solely community based. Three were community based and affiliated with a medical school. The programs represented urban, suburban, and rural practice settings with program sizes ranging from 13 to 38 residents ( $M=22.5$ , standard deviation [SD]=8.7). The number of surveys submitted by each program ranged from 161 to 481 ( $M=322$ ,  $SD=112$ ).

### *The CAT*

The CAT is a 15-item survey that is easily administered in a paper-and-pencil format or via the phone or Internet. A 14-item version for residency programs is also available; it omits an item on whether the doctor's staff treated the patient with respect. Makoul et al published data for practicing physicians across multiple different specialties.<sup>1</sup> These data are essential in allowing users of the tool to benchmark their performance against a national sample. Such comparison data is limited for physicians in training, however, which limits the ability of residents to compare their results on the CAT to those of a broader sample of residents. Wayne et al<sup>11</sup> report pilot data for internal medicine residents as a part of a study examining minimum passing scores for the CAT. For their sample of 30 postgraduate year (PGY)-2 and PGY-3 residents, the average percentage of items rated as "excellent" overall was 71.9% (range=53.2%–87.5%).

The CAT asks respondents to rate different dimensions of the communication and interpersonal skills of the physician using a 5-point rating scale (1=poor, 2=fair, 3=good, 4=very good, 5=excellent). Makoul et al found that scoring the CAT based on the proportion of items rated as excellent was more meaningful than summarizing the scores using means. Their psychometric analysis of response scales indicated that "a rating of 'excellent' was akin to 'yes', while even 'very good' was closer to 'no' than 'yes.'"<sup>11</sup> Data for the present study were analyzed using both means and the percentage of items rated as excellent.

### *Statistics*

For each survey, the mean overall score was calculated as the sum of the item scores divided by the number of items answered. The percentage of items scored as excellent was calculated as the percentage of items with a score of 5 (excellent) out of the number of items answered. The overall mean score and overall percentage of excellent scores were summarized across surveys and stratified by year of training, by native language of the resident, and by gender of the resident.

Because the distributions were not normally distributed, nonparametric tests were performed. For the comparisons by year of training, Kruskal-Wallis one-way analysis of variance tests were used to look for group differences. If there was at least one difference at the  $P<.05$  level, then training levels were compared two at a time using Mann-Whitney U nonparametric tests.

For comparisons by native language of the resident and by gender, Mann-Whitney U tests were used. For the purpose of this study, significant and substantial group differences are defined with a  $P$  value less than .05.

We also examined the percentage of excellent scores for each individual question stratified by training level,

native language of the resident, and gender of the resident. Data are presented as percentage of surveys with an excellent score on each given question. Chi-square tests were used to compare the groups. Again, significant and substantial group differences were defined with a *P* value less than .05.

## Results

Surveys were completed by 1,931 patients. If a patient answered less than 12 of the 14 CAT items, data from that patient were not included in the analysis. This rationale reflected concerns that some unidentified barrier prohibited accurate completion. Seventeen surveys were excluded from the analysis for this reason. An additional 34 surveys were not included because the resident being evaluated was not identified on the form. Data analysis is thus based on 1,880 CAT surveys from patients. Table 1 shows the demographic characteristics of the patients in this sample.

Surveys were collected on 127 residents representing the six family medicine residency programs. Table 2 shows the demographic characteristics of the residents as provided by the residency programs.

### Overall Scores

The overall mean percent for “excellent” was 69.7% (SD=40.4) across items. The overall mean rating on the CAT was 4.59 (SD=0.63). Table 3 shows the mean percentage of the individual CAT items rated as excellent as well as the means and SDs for each item. The medians are also reported due to the highly skewed nature of the data.

The items rated most frequently as excellent were “Paid attention to me (looked at me, listened carefully)” (73.6%), “Treated me with respect” (72.8%), and “Showed care and concern” (72.6%). The items rated least frequently as excellent were, “Encouraged me to ask questions” and “Involved me in decisions as much as I wanted,” 63.2% and 64.9% respectively.

### Differences by Training Year

Table 4 shows the number of surveys completed stratified by PGY. There was substantial variability in the number of surveys submitted per resident based on the resident’s year in training. PGY-1 residents had the fewest number of surveys on average, while PGY-3 residents had the most surveys submitted on average per resident, which reflects their varying numbers of patients per session.

The overall percentage of items scored as excellent varied significantly based on the training year ( $\chi^2=8.38$ , 2 df, *P*=.015). The overall percentage of items scored as excellent was 77.0% (SD=37.4) among PGY-1 residents. This was significantly higher than PGY-2 (M=69.5%, SD=40.2; *z*=-2.38, 233, 626, *P*=.018) and PGY-3 (M=68.1%, SD=41.0, *z*=-2.88, 233, 1021,

Table 1

Demographic Characteristics of Patients Completing the CAT (n=1,880)

	%	<i>n</i>
Gender		
Male	27.2	490
Female	72.8	1,314
Age		
24 or younger	23.4	424
25–44	34.3	620
45–64	31.5	570
65–84	10.1	182
85 or older	0.77	14
Race/ethnicity		
American Indian or Alaska Native	0.95	17
Asian or Asian American	0.84	15
Black or African American	28.1	504
Hispanic or Latino	14.8	266
Native Hawaiian or Pacific Islander	0.06	1
White or Caucasian	51.7	926
Other	3.52	63
Had the patient seen this physician before?		
No	37.1	666
Yes, but only once	15.9	285
Yes, more than once	47.0	842

Note: Surveys were included if the patient did not complete the demographic data on the survey form since none of the planned analyses were based on the demographic characteristics of the patients. Thus, the numbers above will not necessarily total 1,880.

CAT—Communication Assessment Tool

*P*=.004) residents. There were no significant differences between the PGY-2 residents and the PGY-3 residents based on the percentage of items rated as excellent (*z*=-.600, 626, 1021, *P*=.55).

This pattern was also found in the overall means by training year ( $\chi^2=9.95$ , 2 df, *P*=.007). The overall mean ratings for PGY-1 residents (M=4.71, SD=0.55) was statistically significantly higher than the PGY-2 (M=4.59, SD=0.61, *z*=-2.58, 233, 626, *P*=.010) and PGY-3 (M=4.56, SD=0.65; *z*=-3.14, 233, 1021, *P*=.002) residents. There were no significant differences in overall mean ratings between the PGY-2 and PGY-3 residents (*z*=-.671, 626, 1021, *P*=.50).

Table 5 shows the percentage of each CAT item rated as excellent by training year. On nine of the 14 CAT items (64%), the PGY-1 residents were rated as excellent

Table 2  
Demographic Characteristics  
of the Residents (n=127)

	%	n
Gender		
Male	44.1	56
Female	55.9	71
PGY training year		
PGY-1	34.6	44
PGY-2	30.0	38
PGY-3	35.4	45
Native English speaking		
Yes	56.7	72
No	43.3	55

	Mean (SD)	Minimum	Maximum
Age	32.4 (5.8)	25	51

PGY—Postgraduate year  
SD—Standard deviation

more frequently than the PGY-2 and PGY-3 residents. There were no significant differences between PGY-2 and PGY-3 residents on any of the individual CAT items.

#### *Differences by Native Language of the Resident*

Patients completed 861 surveys about residents whose native language was not English (n=55) and 1,019 surveys on native English-speaking residents (n=72). There were no statistically significant differences found in the overall percentage of items rated as excellent when comparing native English-speaking residents (M=71.8%, SD=39.1) and non-native English-speaking residents (M=67.1%, SD=41.7,  $z=-1.86$ , 861, 1,019,  $P=.06$ ). There were, however, statistically significant differences found in the mean overall ratings of the native English-speaking residents (M=4.63, SD=0.60) compared to the non-native English-speaking residents (M=4.54, SD=0.66,  $z=2.09$ , 861, 1,019,  $P=.04$ ).

#### *Difference by Gender of the Resident*

Patients completed surveys about 991 female residents and 889 male residents. There were no significant differences found in the overall percentage of items rated as excellent when comparing female residents (M=70.1%, SD=40.2) and male residents (M=69.2%, SD=40.5;  $z=-.243$ , 991, 889,  $P=.81$ ). There were also no significant differences on the percentage of female

residents versus male residents rated as excellent on any of the individual CAT items. Similarly, there were no significant differences found in the mean overall ratings of female residents (M=4.58, SD=0.65) compared to male residents (M=4.59, SD=0.61,  $z=-.056$ , 991, 889,  $P=.96$ ).

#### **Discussion**

Patient-physician communication is complex and therefore, assessment of communication skills should sample from multiple domains and reporters. The CAT is a “patient experience” measure of the physician-patient relationship and communication. Being able to compare performance of residents against a benchmark group of trainees increases its utility as an evaluation and assessment tool.

Overall, mean ratings on the CAT were high, with respondents rating residents positively. The lowest mean rating on items was 4.47, on a 5-point scale, suggesting the CAT tends to elicit a significant majority of excellent ratings from respondent patients. This ceiling effect is consistent with what is typically seen on patient satisfaction surveys.<sup>12-15</sup> When the data were analyzed based on the percentage of items rated as excellent, however, the ceiling effect is reduced. This reinforces Makoul’s<sup>1</sup> recommendation to score the CAT based on the percentage of items rated as excellent. A measure with a lower ceiling effect is more useful to monitor changes over time and changes due to interventions.<sup>16</sup>

It was anticipated that residents with more training and experience would receive higher patient ratings on the CAT. However, an unexpected finding was that PGY-1 residents were rated higher than PGY-2 and PGY-3 residents. Our sample was accrued in November, which is within the first 5 months of the residency program year. For many programs, first-year residents have considerably more time to work with a patient, typically seeing one patient per hour. Second- and third-year residents are typically scheduled to see four patients per hour. For patients of first-year residents then, this may very well translate into perception of enhanced communication, because there is more time spent with the patient. Interestingly, however, patients did not rate PGY-1s differently than second- or third-year residents on the item “Spent the right amount of time with me.” Thus, while it is plausible that the amount of time that a resident spends with the patient influences the patient’s perception of the resident’s communication skills, how the time is used also impacts the patient’s perceptions. A resident with strong communication skills can facilitate a brief encounter with a patient where the patient still feels that they were heard and their needs were met.

The differences associated with year of training need to be interpreted with caution, however, due to the fact that far fewer surveys were completed about



Table 3  
Percentage of Excellent Ratings and Means for Individual CAT Items

CAT Item	Ratings (% Excellent)	Mean (SD)	Median
1. Greeted me in a way that made me feel comfortable	67.4	4.6 (0.69)	5
2. Treated me with respect	72.8	4.7 (0.63)	5
3. Showed interest in my ideas about my health	69.0	4.6 (0.70)	5
4. Understood my main health concerns	69.5	4.6 (0.70)	5
5. Paid attention to me (looked at me, listened carefully)	73.6	4.7 (0.66)	5
6. Let me talk without interruptions	71.7	4.6 (0.67)	5
7. Gave me as much information as I wanted	69.4	4.6 (0.72)	5
8. Talked in terms I could understand	70.5	4.6 (0.67)	5
9. Checked to be sure I understood everything	68.9	4.6 (0.70)	5
10. Encouraged me to ask questions	63.2	4.5 (0.80)	5
11. Involved me in decisions as much as I wanted	64.9	4.5 (0.78)	5
12. Discussed next steps, including any follow-up plans	71.8	4.6 (0.70)	5
13. Showed care and concern	72.6	4.6 (0.68)	5
14. Spent the right amount of time with me	71.3	4.6 (0.73)	5

CAT—Communication Assessment Tool  
SD—Standard deviation

PGY-1s than about PGY-2s and PGY-3s, which influences the stability of the ratings for PGY-1s. Gathering enough surveys to be able to generalize the findings was a challenge with a 4-week data collection period. Makoul et al<sup>1</sup> recommend gathering a minimum of 20 surveys per resident to have confidence that the results will give a true picture of patient perceptions of that resident's communication skills. Makoul et al base this recommendation on the Rasch generalizability theory, "which estimates that 12–30 ratings per examinee are

required when seeking a reliability of 0.96 for data collected on a 5-point scale."<sup>1</sup> Programs that wish to use the CAT as an assessment or evaluation tool will likely need to collect data for a longer period of time to ensure that first-year residents meet the minimum threshold of 20 surveys.

The native language of the residents did not appear to make a substantial difference with patient perception of overall communication skills as rated using the CAT. While there were significant differences in mean ratings of native English-speaking residents versus non-native English-speaking residents, there were no significant differences based on the overall percentage of items rated as excellent. It should be noted that we dichotomized the language of the resident based solely on whether or not English was his or her native language. We did not include analysis of factors such as the duration of time that the non-native English-speaking residents had spent in the United States, their actual native language, or language in which they received their medical training. Future studies may want to examine these factors more closely.

It was interesting to note that the items rated the highest by patients for residents in the present study matched the highest rated items of practicing physicians as originally published by Makoul et al.<sup>1</sup> The original normative data published on the CAT was based on patient ratings of 38 physicians practicing in a variety

Table 4

Number of Surveys Completed by PGY Year

	PGY-1	PGY-2	PGY-3
Count	44	38	45
Minimum	1	1	1
Maximum	17	44	61
Mean	5.30	16.5	22.7
SD	3.53	10.7	16.0

PGY—Postgraduate year  
SD—Standard deviation

Table 5  
Percentage of Excellent Ratings for Individual CAT Items by Training Year

CAT Item	% Rated as Excellent			Overall P Value	Comparing Groups Two at a Time		
	PGY-1 n=233	PGY-2 n=626	PGY-3 n=1,021		1 Versus 2	1 Versus 3	2 Versus 3
1	75.5%	66.2%	66.2%	.018	*	*	
2	79.4%	73.0%	71.3%	.020	*	*	
3	76.0%	68.9%	67.4%	.040	*	*	
4	74.3%	67.8%	69.4%	.19			
5	79.0%	74.8%	71.7%	.052			
6	80.7%	72.2%	69.4%	.002	*	*	
7	74.5%	69.2%	68.4%	.20			
8	77.3%	70.1%	69.1%	.048	*	*	
9	76.0%	69.9%	66.6%	.016	*	*	
10	74.7%	61.9%	61.4%	<.001	*	*	
11	75.9%	64.3%	62.7%	<.001	*	*	
12	77.6%	72.1%	70.3%	.08			
13	79.6%	71.9%	71.5%	.041	*	*	
14	77.7%	71.4%	69.7%	.053			

\* Significant and substantial group differences are defined with  $P < .05$

Note: See Table 3 for actual CAT item language

CAT—Communication Assessment Tool  
PGY—Postgraduate year

of specialties including dermatology, family medicine, neurosurgery, ophthalmology, orthopedic surgery, and physical medicine and rehabilitation (n=950), all of whom volunteered for the study. While the mean ratings and mean percentage of items rated as excellent were consistently higher for the practicing physician group compared to the resident sample, the items that were rated excellent the most frequently in both groups were “Treated me with respect” (residents=72.8%, physicians=84.4%) and “Paid attention to me (looked at me, listened carefully)” (residents=73.6%, physicians=81.3%). Similarly, both studies found that patients rated the item “Encouraged me to ask questions” as excellent least frequently (residents=63.2%, physicians=62.7%). Other lower-rated items in both groups were “Checked to be sure I understood everything” (residents=68.9%, physicians=70.2%) and “Involved me in decisions as much as I wanted” (residents=64.9%, physicians=70.7%). Thus, it appears that patients perceive that both physicians in training and practicing physicians are respectful and pay attention to them. However, patients are less satisfied with their active involvement in the

decision-making process regarding their care and their opportunities to ask questions during the encounter.

#### Limitations

Limitations of this study include the fact that the family medicine residency programs from which data were collected were selected as a convenience sample of volunteering programs. While participating programs represented a diversity of settings and populations served, they do not represent the scope of family medicine residency programs across the United States.

Further, our sample represents only the beginning of some initial benchmarking data for the use of the CAT in family medicine residency programs. To truly develop a normative data sample for the CAT, future studies will want to substantially increase the number of participating programs.

Additionally, there was substantial variability in the number of surveys gathered per resident, which limits the stability and generalizability of the data. Future studies may want to extend the data collection time to ensure that a minimum of 20 surveys are collected for each resident included in the normative sample.

## Conclusions

The CAT is an effective and efficient tool for assessing patient perceptions of physicians' communication skills. In a busy residency program, the CAT is easily administered, scored, and understood. Scoring the CAT based on the percentage of items that are rated as excellent clearly differentiates residents and highlights the strengths and weakness for each resident. We recommend that the results of the CAT be utilized as both an evaluative and learning tool.

As an evaluative tool, we recommend that the results of the CAT be reported to each resident to provide systematic feedback that can increase awareness of patients' perceptions regarding their communication skills. By using benchmark data (ie, comparing the resident to others in their program and to the present sample), residents and program directors have an objective measure of interpersonal and communication skills that can be used for ongoing evaluation of this critical ACGME core competency. Programs can also identify areas of strength and weaknesses that may need to be addressed further in their curriculum.

As a learning tool, the results of the CAT can be presented to the resident either individually or in a group setting, with a focus on how the CAT results can assist them in improving their communication skills. Along with a thorough explanation of the results of the CAT, it is recommended that a faculty member guide the residents in identifying their strengths (top three scores) and areas of growth (lowest three scores). Some items may receive low ratings across all residents in the program, signaling an opportunity for group training, while more idiosyncratic areas for improvement will be tailored to individuals. Performance improvement goals with an action plan should be developed related to the residents' areas of growth. The CAT could easily be repeated biannually or annually to provide feedback to the residents on their progress toward individual goals.

Another learning option for the CAT focuses on self-assessment. Each resident could complete the CAT for themselves based on how they think their patients would rank them before obtaining their individualized results. Residents could identify what items they believe are their communication strengths and weaknesses. This self-assessment could then be compared to the actual results to determine if their perception of their communication skills matches their patients' perception. Using the CAT in this manner may lead to even greater mindfulness in the area of interpersonal and communication skills.<sup>17</sup>

*Acknowledgments:* Ana Rivera-Tovar, PhD, translated the CAT into Spanish. The Spanish version is available from the corresponding author or from Gregory Makoul, PhD (GMakoul@stfranciscare.org).

Thank you to Nancy Buderer for assistance with the statistical analysis and to Janhangir Adil, MD, for programmatic support of this project.

This paper was presented at the International Conference on Communication in Healthcare, October 6, 2009, in Miami.

*Corresponding Author:* Address correspondence to Dr Myerholtz, Mercy Health Partners, Mercy Family Medicine Residency Program, 2200 Jefferson Avenue, Toledo, OH 43604. 419-251-1819. Fax: 419-242-9806. linda\_myerholtz@mhsnr.org.

## REFERENCES

1. Makoul G, Krupat E, Chang C. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. *Patient Educ Couns* 2007;67:333-42.
2. Accreditation Council for Graduate Medical Education. The Common Program Requirements 2007. Available from [www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf](http://www.acgme.org/outcome/comp/GeneralCompetenciesStandards21307.pdf). Accessed March 6, 2009.
3. Duffy F, Gordon G, Whelan G, Cole-Kelly K, Frankel R, all participants in the American Academy on Physician and Patient's Conference on Education and Evaluation of Competence in Communication and Interpersonal Skills. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. *Acad Med* 2004;79:495-507.
4. Lang F, McCord R, Harvill L, Anderson D. Communication assessment using the common ground instrument: psychometric properties. *Fam Med* 2004;36(3):189-98.
5. Schirmer JM, Mauksch L, Lang F, et al. Assessing communication competence: a review of current tools. *Fam Med* 2005;37(3):184-92.
6. Swing S. Assessing the ACGME General Competencies; general considerations and assessment methods. *Acad Emerg Med* 2002;9:1278-88.
7. 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.
8. Lynch D, Swing S. ACGME Outcome Project: key considerations for selecting assessment instruments and implementing assessment systems. [www.acgme.org/outcome/assess/keyConsider.asp](http://www.acgme.org/outcome/assess/keyConsider.asp). Accessed August 12, 2009.
9. Epstein R, Hundert E. Defining and assessing professional competence. *JAMA* 2002;287:226-35.
10. Sitzia J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int J Qual Health Care* 1999;11(4):319-28.
11. Wayne D, Cohen E, Makoul G, McGaghie W. The impact of judge selection on standard setting for a patient survey of physician communication skills. *Acad Med* 2008;83:S17-S20.
12. Moret L, Nguyen JM, Pillet N, Falissard B, Lombrail P, Gasquet I. Improvement of psychometric properties of a scale measuring inpatient satisfaction with care: a better response rate and a reduction of the ceiling effect. *BMC Health Serv Res* 2007;7:197-206.
13. Pouwer F, Snoek F, Heine R. Ceiling effect reduces the validity of the Diabetes Treatment Satisfaction Questionnaire. *Diabetes Care* 1998; 21:2039.
14. Salisbury C, Burgess A, Lattimer V, et al. Developing a standard short questionnaire for the assessment of patient satisfaction with out-of-hours primary care. *Fam Pract* 2005;22:560-9.
15. Vedsted P, Sokolowski I, Heje N. Data quality and confirmatory factor analysis of the Danish EUROPEP questionnaire on patient evaluation of general practice. *Scand J Prim Health Care* 2008; 26:174-80.
16. Kimberlin C, Winterstein A. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008;65:2276-84.
17. Epstein RM. Mindful practice. *JAMA* 1999;282:833-9.