The resident view of professionalism behavior frequency in outstanding and “not outstanding” faculty

Kimberly Ephgrave, M.D., F.A.C.S. a,*, R. Brent Stansfield, Ph.D. b, Jerold Woodhead, M.D. c, W. John Sharp, M.D., F.A.C.S. a, Thomas George, M.D. c, John Lawrence, M.D., F.A.C.S. a

a Department of Surgery, University of Iowa Hospitals and Clinics, 200 Hawkins Dr, Iowa City, IA 52242-1009, United States
b Office for Consultation and Research in Medical Education, University of Iowa Carver College of Medicine, 1216 Medical Education Building, Iowa City, IA 52242-1009, United States
c Department of Pediatrics, University of Iowa Hospitals and Clinics, 200 Hawkins Dr, Iowa City, IA 52242-1009, United States

Manuscript received April 7, 2005; revised manuscript February 10, 2006

Abstract

Background: Professionalism assessment has become necessary for all postgraduate training programs because it is now required for accreditation. To validate the novel items we generated to assess professionalism, we tested whether residents’ ratings of faculty they judged as outstanding in professionalism would be distinguishable from those they judged as not outstanding.

Methods: Educators from core clinical disciplines generated 20 items assessing professionalism behaviors on a 7-point frequency scale anchored by “always” and “never.” Thirty-five surgical and pediatric residents completed the form twice, anonymously rating 1 faculty member they judged as outstanding and another they judged as not outstanding.

Results: The residents produced 69 faculty ratings with means that differed significantly on all items between the outstanding and not-outstanding faculty. The form was highly unidimensional, with the primary factor’s eigenvalue being 11.5 and Cronbach’s alpha being 0.97. Groups differed most on items, ie, “listens well,” “inspires trust,” “answers questions directly,” and “demonstrates respect for all.”

Conclusion: The behaviors that best distinguished clinical faculty judged by residents as outstanding professionals were listening, trustworthiness, answering directly, and respect. © 2006 Excerpta Medica Inc. All rights reserved.

Keywords: Accreditation Council for Graduate Medical Education competencies; Assessment; Evaluation; Pediatrics; Professionalism; Surgery

Professionalism discussions have been growing in medical education for the last several decades [1]. All postgraduate training programs must currently identify or develop usable tools to assess professionalism in their trainees because professionalism is 1 of the 6 competencies that must be assessed for postgraduate program accreditation by the Accreditation Council for Graduate Medical Education.

For medical students, clinical training is not known for enhancing professionalism. Unfortunately, past research suggests that students’ clinical training is associated with a decrease in empathy and increase in cynicism compared with the attitudes students hold when they begin medical school [2–4]. This attitude deterioration in medical students may be a consequence of residents and faculty purveying a “hidden curriculum,” which can undermine schools’ explicit teachings of ethics and professional values [5]. Specific analysis of the recommended versus the actually taught curriculum in 1 internal medicine clerkship, for instance, found that although both interprofessional respect and the importance of service were present in the explicitly taught curriculum, the actual ward experiences that were observed, recorded, and analyzed by the study team more frequently demonstrated interprofessional disrespect and the burden of service [6].

Residents and students were included in a 2002 invitational conference, “Embedding Professionalism in Medical Education: Assessment As a Tool for Implementation,” cosponsored by the Association of American Medical Colleges (AAMC) and the National Board of Medical Examiners (NBME). The resident and student participants emphasized the importance of faculty role modeling in their learning environment, stating “We’re being asked to be pro-
professional in an unprofessional environment. Faculty should be subject to the same criteria for assessment as students and residents.” [7]. Residents at our own institution echoed this sentiment.

The purpose of the current work was 3 fold. First, we sought to measure professionalism above the medical student level by comparing residents’ responses on those items regarding faculty they perceived as outstanding versus those whom they perceived as not outstanding in professionalism. Second, we wished to determine which behaviors were significantly more frequent in the outstanding compared with the not-outstanding cohort. The consensus of our group of educators from 4 core clinical disciplines was that residents were our primary concern because they had been relatively neglected as a focus of medical education until the adoption of the competency requirement. Assessing the role modeling of professionalism from faculty was also an important goal because of the impact of faculty behaviors on both medical student and resident behavior. Our professionalism task force therefore worked to create a tool by which residents could rate the frequency of professionalism behaviors of their faculty and be rated in turn on their behavioral frequencies by the faculty. Our final goal was to develop a very brief professionalism assessment tool (i.e., one that would easily fit on a single printed page or single computer screen) because both residents and faculty suffer from rater fatigue.

We tested the form by asking residents to anonymously rate unidentified faculty members. We found that our residents concurred with the opinion of residents and students at the national AAMC/NBME-sponsored conference on professionalism that faculty should be subject to the same professionalism assessment as trainees. Thus, our residents were relatively eager to participate in this opportunity to try out our professionalism assessment instrument on faculty.

Methods

Leaders in medical education including both clerkship and residency program directors from surgery, pediatrics, internal medicine, and obstetrics and gynecology joined the associate dean for Student Affairs And Curriculum and staff from the Office for Consultation and Research in Medical Education to develop a tool to rate the frequency of professionalism behaviors that residents and faculty could observe in each other. We used as resources the December 2003 Request for Proposals put out by the Center for Innovation of the NBME, which included the proceedings of their recent national invitational conference on professionalism in medical education, which as cosponsored by the AAMC. These materials included specific positive and negative behaviors associated with 8 categories of professionalism: altruism, honor and integrity, caring, compassion and communication, respect, responsibility, accountability, excellence and scholarship, and leadership. The categories have some overlap with the 6 positive and 6 negative attributes of professionalism put forward several years earlier by the Professionalism Project of the American Board of Internal Medicine, which in turn derived in part from their Humanism Project in the 1980s.

We concentrated on 5 categories: altruism; honor and integrity; respect; caring, compassion, and communication; and responsibility and accountability. We generated 20 novel items stated as behaviors that could be rated on a 7-point frequency scale anchored by “always” and “never.” We used both positive and negative behaviors and reverse-scored the negative behaviors so that high scores were always indicative of a higher level of professionalism.

Educational leaders in pediatrics and surgery took rating forms to meetings with their residents and asked the residents to anonymously fill out the forms for 2 faculty members they knew well: 1 whom they judged outstanding in professionalism, and 1 whom they judged to be not outstanding. No record was made of which residents filled out the forms, and the form did not ask for identification of the faculty members that the residents had in mind for their ratings. This study therefore had no identifiers for either the raters or the raters and therefore was given exempt status by the Human Subjects Institutional Review Board of the University of Iowa.

We obtained 32 faculty ratings from 16 surgical residents, and 37 faculty ratings from 19 pediatrics residents, for a total of 35 outstanding faculty ratings and 34 not-outstanding faculty ratings. (One pediatric resident reported not knowing a faculty member who was not outstanding in professionalism well enough to rank.)

Mean scores across the 20 items were compared between groups (outstanding vs. not-outstanding faculty) using paired-sample Student \( t \) test. Post hoc paired-sample Student \( t \) tests were conducted separately for each item using Scheffe correction for family-wise error. Effect size for each item between groups was interpreted as construct validity of that item.

Each rater’s average rating on negatively worded (reverse-coded) items and average rating on positively worded items were compared to test whether those items were systematically less informative because of item misinterpretation. Three paired-sample Student \( t \) tests were employed: 1 across all rated faculty (testing whether item type affected overall rating bias) and 2 separately within each faculty group.

The 20-item form was analyzed as a whole with Cronbach’s alpha to estimate its reliability and a principle component factor analysis to measure its unidimensionality. Cronbach’s alpha was computed separately for each faculty group and for all (outstanding and not outstanding) faculty together. The factor analysis was performed using SAS version 8.0 (proc FACTOR) using all rated faculty and ignoring rater identity. To produce a very short measurement instrument with similar reliability, the 10 items loading highest on the primary factor (items no.1, 3, 4, 9, 13, 14, 15, 16, 17, and 20; Table 1) were selected. Cronbach’s alpha was computed for this subset to estimate the reliability of this instrument relative to that of the entire 20-item form.
Table 1

<table>
<thead>
<tr>
<th>Item nos.</th>
<th>Survey items</th>
<th>Outstanding Mean (95% CI)</th>
<th>Not outstanding Mean (95% CI)</th>
<th>t(33) =</th>
<th>P values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Demonstrates respect toward all others both in direct interactions and in indirect references</td>
<td>6.3 ± 0.8 3.3 ± 1.6</td>
<td>2.94 (2.24–3.64)</td>
<td>8.58</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2.</td>
<td>Is prepared for clinical responsibilities</td>
<td>6.3 ± 0.8 4.6 ± 1.6</td>
<td>1.76 (1.17–2.36)</td>
<td>6.03</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3.</td>
<td>Understands and attends to the learner’s needs</td>
<td>6.0 ± 0.9 3.4 ± 1.5</td>
<td>2.62 (1.99–3.24)</td>
<td>8.52</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4.</td>
<td>Loses control over emotions, e.g. swearing, yelling, or throwing objects*</td>
<td>6.6 ± 1.0 4.1 ± 1.8</td>
<td>2.44 (1.71–3.18)</td>
<td>6.76</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>5.</td>
<td>Responds to requests for assistance, including pages and e-mail, in a timely and supportive manner</td>
<td>6.0 ± 0.8 3.9 ± 1.6</td>
<td>2.00 (1.17–2.83)</td>
<td>4.93</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>6.</td>
<td>Is insensitive to patient’s values and beliefs*</td>
<td>5.6 ± 2.0 3.8 ± 1.7</td>
<td>1.39 (0.28–2.51)</td>
<td>2.54</td>
<td>.061</td>
</tr>
<tr>
<td>7.</td>
<td>Aware of own limitations; seeks and accepts constructive feedback</td>
<td>5.4 ± 1.3 2.8 ± 1.5</td>
<td>2.33 (1.57–3.09)</td>
<td>6.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>8.</td>
<td>Promotes own interests above the common good of teams or work groups*</td>
<td>5.4 ± 1.7 4.0 ± 1.8</td>
<td>1.33 (0.36–2.30)</td>
<td>2.80</td>
<td>.086</td>
</tr>
<tr>
<td>9.</td>
<td>Answers questions directly and respectfully</td>
<td>6.5 ± 0.6 3.5 ± 1.5</td>
<td>3.06 (2.45–3.67)</td>
<td>10.26</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>10.</td>
<td>Deals with confidential information discretely and appropriately</td>
<td>6.0 ± 1.0 3.8 ± 1.7</td>
<td>1.88 (1.06–2.70)</td>
<td>4.65</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>11.</td>
<td>Puts own needs ahead of patients’ needs*</td>
<td>5.5 ± 1.7 4.5 ± 1.7</td>
<td>.82 (–0.09–1.74)</td>
<td>1.82</td>
<td>.0774</td>
</tr>
<tr>
<td>12.</td>
<td>Tactfully offers assistance and support for team members</td>
<td>5.9 ± .8 2.5 ± 1.2</td>
<td>3.18 (2.43–3.94)</td>
<td>8.59</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>13.</td>
<td>Is unfair in making decisions affecting coworkers or subordinates*</td>
<td>6.4 ± 1.0 3.8 ± 1.8</td>
<td>2.24 (1.42–3.06)</td>
<td>5.55</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>14.</td>
<td>Inspires trust in patients, colleagues, coworkers and subordinates</td>
<td>6.2 ± 1.1 3.0 ± 1.2</td>
<td>3.03 (2.40–3.66)</td>
<td>9.72</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>15.</td>
<td>Is prepared for teaching responsibilities</td>
<td>6.2 ± 1.3 4.0 ± 1.8</td>
<td>2.30 (1.55–3.06)</td>
<td>6.22</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>16.</td>
<td>Listens well and responds appropriately</td>
<td>6.4 ± 0.7 3.1 ± 1.5</td>
<td>3.27 (2.66–3.88)</td>
<td>10.94</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>17.</td>
<td>Gives colleagues due credit</td>
<td>6.1 ± 0.9 3.3 ± 1.7</td>
<td>3.18 (2.54–3.82)</td>
<td>10.10</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>18.</td>
<td>Inappropriately crosses personal boundaries with patients or coworkers*</td>
<td>6.3 ± 1.2 4.4 ± 1.8</td>
<td>1.91 (1.24–2.58)</td>
<td>5.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>19.</td>
<td>Demonstrates personal integrity and a willingness to take a stand on principle</td>
<td>6.3 ± .9 3.5 ± 1.7</td>
<td>2.55 (1.65–3.44)</td>
<td>5.82</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>20.</td>
<td>Exemplifies professional behavior</td>
<td>6.3 ± .8 2.9 ± 1.3</td>
<td>3.42 (2.85–4.00)</td>
<td>12.14</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

CI = confidence interval.
* Reversed score.

Results

The outstanding group’s mean scores were greater than those of the not-outstanding for all items. The magnitude of that difference ranged from a low of 1.0 (“puts own needs ahead of patients’ needs”) to a high of 3.3 (“listens well and responds appropriately”) on the 7-point frequency scale (see Table 1 listing all items and means and variance for both groups). Means differed significantly for 19 of the 20 items ($t(33) = 12.10$, $P < .0001$).

Most of the 20 items demonstrated near full use of the 7-point scale for the not-outstanding group, whereas the outstanding distributions were positively skewed and less variable. For instance, the item “gives colleagues due credit” (see Fig. 1) had a mean of 3.3 and SD of 1.7 for the not-outstanding group compared with 6.1 and 0.9, respectively, for the outstanding group (see Fig. 1).

Items that best distinguished the 2 groups were “listens well and responds appropriately,” “inspires trust in patients, colleagues, coworkers, and subordinates,” “answers questions directly and respectfully,” and “demonstrates respect toward all others both in direct interactions and in indirect references.” All of these items produced group means that differed by at least 3 points on the 7-point frequency scale.

For the outstanding faculty, item means ranged from a low of 5.4 (“aware of own limitations; seeks and accepts constructive feedback” and “promotes own interests above the common good of team or work group”) to a high of 6.6 (“loses control over emotions, e.g., swearing, yelling, or throwing objects”). Because negative behaviors were reverse-coded, the most professional score, i.e., 7, remains the highest number for all 20 items.

For the not-outstanding faculty, item means ranged from 2.5 to 4.6. The least frequent behaviors for the not-outstanding faculty related to relationships with other professionals, i.e., “tactfully offers assistance and support for team members” and “aware of own limitations; seeks and accepts constructive feedback.” The most frequent behaviors for not-outstanding faculty, indicating that these professionalism behaviors were rated to occur more often than not, dealt with patient interactions, i.e., “is prepared for clinical responsibilities” (mean 4.6) and “puts own needs ahead of patients’ needs” (mean 4.5). Thus, the not-outstanding faculty were observed to behave professionally less frequently...
in interactions with other members of the health care team than during patient care.

Whenever negatively worded items are used, misinterpretation of the item is a concern. The average of each rater’s negatively worded item ratings (mean average rating across all faculty was 4.96) and the average of positively worded item ratings (mean average rating across all faculty was 4.62) were not significantly different ($t(68) = 1.93$ [not significant]) across faculty groups. However, average ratings within item type did differ within faculty type: Outstanding faculty ratings were higher on positively worded than negatively worded items ($t(34) = .73$, $P < .001$), and not-outstanding faculty were lower on positively worded than negatively worded items ($t(33) = 5.77$, $P < .0001$). This suggests to us that some item misinterpretation occurred.

The reliability (Cronbach’s alpha) for the full 20-item instrument was 0.97. Shortening the length by using the 10 best discriminating items (items no. 1, 3, 4, 9, 13, 14, 15, 16, 17, and 20) preserved this level of reliability. Factor analysis showed that despite deliberate inclusion of a spectrum of items across 5 categories of professionalism behaviors, the tool was essentially unidimensional. One factor with an eigenvalue of 11.5 explained 57% of the variance. Two additional factors had eigenvalues >1.0 (1.41 and 1.04) but together explained only an additional 13% of the variance. These 2 small factors did not fit any clearly definable construct.

Comments

Previous assessments of professionalism include student and resident [8] assessment of professionalism behaviors in internal medicine residents, faculty assessments of professionalism in surgical residents [9,10], and faculty global assessments of all specialties of residents, all of which were intended to cover each of the 6 recently adopted graduate medical education competencies [11]. Our work complements these previous studies of medical students and residents by showing that a highly reliable professionalism instrument can be developed using resident ratings of faculty behavior.

Clinical-performance evaluation instruments have been most commonly studied in medical students, with factor analyses reported beginning in the early 1960s [12]. More recently studies of resident clinical performance rating instruments appear in the literature and support the findings for medical student clinical-performance rating. A recent large study by Silber et al [11] analyzed faculty ratings of 1295 residents at 2 medical centers in Philadelphia. Similar to multiple previous findings for medical students, Silber et al demonstrated 2 factors rather than the 6 competencies their instrument was designed to assess. The 2 factors that have commonly emerged from global rating forms in past medical student studies are an interpersonal communication and professionalism factor and a medical knowledge and clinical skills factor. Together those factors accounted for 77% of the variance in Silber et al’s resident ratings. The professionalism factor had a much larger eigenvalue of 16.5, compared with an eigenvalue of 1.32 for their medical knowledge and clinical skills factor. In terms of the Accreditation Council for Graduate Medical Education competencies, interpersonal communication and professionalism fed into the factor with the eigenvalue of 16.5, whereas the other 4 competencies contributed to the smaller factor.

The present work is congruent with the previous literature on clinical performance ratings. Despite selecting items culled from different professionalism behaviors—classified as altruism; honor and integrity; caring, compassion, and communication; respect; or responsibility and accountability—our study was similar to others in that essentially 1 strong factor emerged.

Other approaches to assessment of professionalism include simulations with standardized patients, either in a known examination setting or along with a clinician’s actual clinical activity. These have value especially for simulated patients’ judgments of ability to establish patient rapport. Ability to establish rapport may be affected by demographic or personality variables of the simulated patient and examinee, but scores do also improve during the course of medical school [13]. Simulations used for primary care faculty and residents, when unannounced standardized patients are mixed into real clinic settings, are usually not detected as simulations and thus have very high face validity (but unfortunately also high expense). We expect that the expediency, low expense, and reliability of paper or electronic rating forms will necessitate their continued use for the assessment of professionalism whether or not periodic simulations are also used.

The present study shows that the behaviors differing most between faculty considered by residents as outstanding or not outstanding in professionalism were “tactfully offers assistance and support for team members,” “listens well and responds appropriately,” “inspires trust in patients, colleagues, coworkers, and subordinates,” “answers questions directly and respectfully,” and “demonstrates respect toward all others both in direct interactions and in indirect references.” These items seem important and represent several different
aspects of professionalism (i.e., respect; honor and integrity; and caring, compassion, communication). However, these data are from a preliminary study and therefore have significant limitations. The data were collected on the most extreme faculty members the residents knew well enough to rate, so the best discriminating items may change when all real faculty and residents in a clinical department rate each other. We do not know, for instance, if many residents in this preliminary study may have evaluated the same few outstanding faculty members.

In the immediate future, having the same instrument used across the continuum from first-year residents to senior faculty will allow us to develop norms for professionalism and expectations for improvement with various amounts of clinical experience. Using the instrument in several disciplines with different activities and cultures will also allow us to determine if the mental construct of professionalism or frequencies of behaviors differ between disciplines. Finally, it is reasonable to hope that constructively communicated formative feedback from this instrument may accelerate the improvement in professionalism that occurs with clinical experience.

References


