

## Using Standardized Patients to Assess Medical Students' Professionalism

MICHAEL D. PRISLIN, DÉSORÉE LIE, JOHANNA SHAPIRO, JOHN BOKER, and STEPHEN RADECKI

The subject of professionalism is currently engendering great interest within the medical education community. Concern exists that conditions within the health care delivery environment threaten established standards of professional behavior, and, perhaps more insidiously, that the medical education experience itself may be negatively influencing the development of physicians' professionalism.<sup>1-3</sup> As a consequence, much energy has recently been directed toward defining competencies that reflect professionalism and in creating corresponding curricula that will foster learning in this domain.<sup>4-6</sup>

However, having instruments that can accurately measure the attainment of professionalism remains an elusive goal.<sup>7-9</sup> This study examines the utility of standardized patient-based assessments of professional characteristics. Comparisons are made with other measures of professionalism, such as faculty evaluation, performance on a written self-reflective exercise, and student-reported participation in community service activities.

### Method

This study was conducted at the University of California, Irvine (UCI), College of Medicine. Participants were students completing the year two patient-doctor course during the 1999-00 academic year. This course represents the second segment of a vertically integrated four-year course sequence in professional skill development. The year two segment focuses on patient-physician communications, physical diagnosis, and the development of basic clinical reasoning skills. Eight core clinical modules are linked to topics concurrently being taught in the year-two pathology, pathophysiology, and pharmacology courses. Each module begins with a standardized patient interaction, followed by generation of learning issues within small tutorial groups. Mid-module activities include topical didactic presentations and physical diagnosis instruction. Each module concludes with a wrap-up session in which the diverse learning activities are tied together through small-group discussion of the original learning issues. These discussions typically feature a heavy emphasis on patient-physician communication and professional behavior.

Assessments of students occurring during the course consist of a written final examination, structured written evaluations completed by the faculty group leaders, and an appraisal of clinical skills. The clinical skills appraisal for 1999-00 consisted of a three-station standardized-patient-based examination. The cases were a patient presenting with fatigue, a patient presenting with upper gastrointestinal and chest discomfort, and a patient presenting with transient neurologic deficits. The first two cases each entailed 25 minutes and the third case entailed 35 minutes of patient contact. Each station required students to perform a history and physical examination. In addition, students performed a rapid computer-based literature search following the initial encounter with the neurology case, written and oral clinical presentations following the fatigue case, and a written reflective essay, pertaining to students' reactions to a poem describing a 39-year-old man experiencing an acute myocardial infarction and sudden death, following the upper gastrointestinal and chest pain case. Each standardized patient encounter included assessments of history and physical exam performance based on a checklist and assessments of communication

skills and professionalism using a rating scale. The rating scale for communication skills used in this study was a modification of the Communications Skills Form developed at East Tennessee State University by Forrest Lang, MD, to assess patient-centered communications as evaluated by standardized patients. It is based upon an instrument developed by the American Board of Internal Medicine to assess patients' satisfaction. The rating scale includes six items relating to communication that are reported here as the cumulative communication score; a single item relating to overall professional competence; and a single item relating to overall standardized patients' satisfaction. The professionalism scale used for this study was constructed based upon the work of Arnold and colleagues,<sup>9</sup> and consisted of three items: one that allowed standardized patients to rate students' knowledge and competence, one that rated students' integrity, and one that rated students' altruism. Taken together, these three items are reported as the cumulative professionalism score. The specific rating scale items for communication and professionalism are presented in List 1.

Both the communication and the professionalism rating instruments used five-point Likert scales with the following specific anchors: 5—outstanding; 4—very good; 3—good; 2—needs improvement; 1—marginal. Therefore, the maximum achievable scores were: cumulative communications—30 points, cumulative professionalism—15 points, professional competence—5 points, and overall satisfaction—5 points. Standardized patients received detailed verbal and written instructions on how to complete the communication scale, including descriptive anchors for performance at varying levels of competence, and were observed rating performances using practice tapes before participating in the examination. In terms of the specific professionalism items, the standardized patients were instructed to respond based upon their own personal perceptions of the students. Fourteen standardized patients were used during the course of the examination: seven for the fatigue case, three for the chest-discomfort case, and four for the neurology case.

Faculty evaluations of students' performances during the patient-doctor II course were based on an 11-item rating scale in which one item assessed whether the student "demonstrates professional behavior." This evaluation also used a five-point Likert scale in which five represented outstanding, four represented above expected, three represented at expected, two represented below expected, and one represented problematic performance. Hence the maximum possible score for faculty professionalism ratings was five points. Evaluating faculty received verbal instructions regarding evaluating students' performances during faculty development sessions. Evaluation of the professionalism item focused on students' citizenship and academic honesty, team participation, and interactions with standardized patients during the interview sessions.

The essay was scored by one of the study's authors for emotional content and problem-solving capacity using a modification of a method described by Pennebaker and colleagues.<sup>10</sup> Subscale scores relating to empathy and positive coping attitudes were used as measures reflecting students' expressions of professional attributes. The scores students received represented a sum of these two subscales.

Students' descriptions of their participation in community service activities were elicited by means of a written survey distributed at the conclusion of the skills-appraisal exercise. Participation was scored as "did" or "did not" participate.

### List 1. Rating-scale Items for Communication and Professionalism

Identified all of my concerns*
Effectively gathered information relevant to my case*
Listened actively*
Demonstrated interest in me as a person*
Addressed my feelings*
Worked to reach common ground*
Appeared professionally competent
My overall satisfaction†
Student appeared to know what he/she was doing‡,§ (knowledgeable and competent)
I trust this student§
This student placed my interest first§

\*Cumulative communications items; †rates overall professional competence; ‡rates overall standardized patient's satisfaction; §cumulative professionalism items.

All standardized patient encounters were videotaped, and 15 were randomly selected from each of the three cases for blind review and scoring on the communication and professionalism rating scales. These independent ratings were undertaken by either one of the study's authors or one of the UCI standardized patient trainers. Inter-rater reliabilities were then assessed using intraclass correlation coefficients calculated through analysis of variance.

Scores of individual students across the three cases for the cumulative communication and cumulative professionalism rating scales were also examined for their relative correlations by using intraclass correlation coefficients calculated from analysis of variance. The contributions of discrete rating scale items in accounting for variations in individual students' scores found across the three cases for the communications were then assessed using a bivariate and stepwise factor analysis.

Finally, all potential correlations of students' overall performances as assessed by cumulative communication scores, cumulative professionalism scores, overall professionalism scores, overall standardized patient satisfaction, faculty professionalism evaluations, reflective essay scores, and participation in community service activities were ascertained using Spearman's rho tests.

### Results

Eighty-five students completed the clinical skills-appraisal exercise, 78 students completed scorable essays, and 54 students completed the community service activity survey. As previously noted, 45 individual clinical skills-appraisal station exercises were randomly selected to assess inter-rater reliability.

The mean SP-derived cumulative communication scores for the three cases were 20.00 (range 11–30) for case 1, 20.32 (range 14–30) for case 2, and 19.74 (range 9–30) for case 3. The mean SP-

derived cumulative professionalism scores were 9.81 (range 5–15), 10.13 (range 6–15), and 9.75 (range 5–15), respectively. The inter-rater reliability (intraclass correlation coefficient) for the cumulative communication scores was .61. For the cumulative professionalism scores the inter-rater reliability (intraclass correlation coefficient) was .65. However, inter-case correlations (intraclass correlation coefficients) of individual students' cumulative communication scores and cumulative professionalism scores were only .16 and .19, respectively. The factor analysis indicated that no one of the six individual communication items or the three individual professionalism items accounted for a disproportionate amount of the inter-case variation.

Correlations between the various SP-derived communication, professionalism, and satisfaction scores and faculty evaluations of professionalism, students' essay performances, and student-reported community service are presented in Table 1. Only the SP-based communication, professionalism, and satisfaction scores demonstrate significant correlation.

### Discussion

This study reports data derived from a single class of students at one medical school in the early stages of their clinical education. Furthermore, the exercise used to derive the SP-based data was, at best, of modest rigor, as defined by the small number of clinical cases and total patient contact time of only 85 minutes and observed inter-rater reliability scores in the .60–.65 range. Thus, this study should be viewed as exploratory.

Nevertheless, we find the results interesting. We were surprised to find a lack of inter-case correlation in both communication and professionalism scores. Hodges and colleagues have reported similar inter-case variation with respect to standardized patients' assessment of communication skills.<sup>11</sup> However, the cases used in their study contained high levels of psychosocial complexity, which were likely to provoke considerable performance variations among the examinees in their study. Given the broad clinical nature and the relative lack of psychosocial complexity in the cases used in our study, one might expect the rating scales to produce assessments of communication and professionalism that would generalize across cases. Or if not, one might at least expect that scores would correlate most strongly with standardized patients' perceptions of students' knowledge as assessed in the context of each specific case. But our data do not demonstrate this. Interestingly, Donnelly and colleagues recently reported similar results with respect to assessing residents' communication skills.<sup>12</sup>

Clearly, SP-based assessments of communication skills, professionalism, and overall satisfaction overlap to a great extent. Indeed, our findings suggest that there is little benefit in having a separate scale for professionalism. Yet, what exactly the standardized patients assessed in this study remains unclear. Arnold and colleagues ana-

TABLE 1. Correlations of Various Measures of Communication and Professionalism

	Cumulative Professionalism	Overall Professional Competence	Overall Satisfaction	Faculty Professionalism Evaluation	Essay Score	Community Service Participation
Cumulative communications	.736†	.661†	.807†	.149	.057	.252
Cumulative professionalism		.804†	.819†	.129	-.009	.222
Overall professional competence			.807†	.118	-.015	.172
Overall satisfaction				.081	-.087	.242
Faculty evaluation					.129	.067
Essay						-.159
Community service participation						

\*Spearman's rho.  
† $p < .01$ .

lyzed 12 items describing medical professionalism as defined by the American Board of Internal Medicine.<sup>9</sup> Their rating scale demonstrated moderately high internal reliability. Furthermore, a factor analysis identified three characteristics, excellence (knowledge/competence), honor/integrity, and altruism/respect, which accounted for most of the variation observed in ratings of medical student and resident professionalism. Hence, we believe that our items also possess at least *prima facie* criterion validity. Yet, our data indicate that standardized patients' assessments of professionalism are distinct from a variety of other measures obtained from non-SP-based observations.

Conceptually, one might view the evolution of medical students' professionalism as a combination of intrinsically held values and externally imposed normative standards. Measuring students' professionalism in turn involves an evaluation of students' behaviors. In this study, these measurements reflect interpretations of students' behaviors undertaken by standardized patients and by faculty. The measurements also include interpretations of implicit behaviors as suggested in the students' self-reflection exercises, and explicit behaviors as manifested by community service participation. Although it seems quite likely that each of these approaches portrays a different perspective of professionalism, the lack of correlation between these measures is striking.

In summary, our data suggest that broad SP measurements of professionalism based on rating scales yield a diffuse assessment of professional characteristics, and that these assessments may, in fact, vary from case to case. Recently, Ginsberg and colleagues proposed that creating specific contextual situations involving conflict between competing professional values might afford a better opportunity to assess professionalism based upon demonstrable behaviors.<sup>13</sup> Further study will be required to determine whether more specific, focused standardized patient case scenarios that necessitate students' management of situations involving conflict of professional values might yield more precise measurements of this crucial medical education competency.

Correspondence: Michael D. Prislun, MD, University of California, Irvine, Department of Family Medicine, Bldg. 200, Rt. 81, Room 512, 101 The City Drive South, Orange, CA 92868; e-mail: {mdprisli@uci.edu}.

---

#### References

1. Hensel WA, Dickey NW. Teaching professionalism: passing the torch. *Acad Med.* 1998;73:865-70.
2. Feudtner C, Christakis DA, Christakis NA. Do clinical clerks suffer ethical erosion? Students' perceptions of their ethical environment and personal development. *Acad Med.* 1994;69:670-9.
3. Hundert EM, Hafferty F, Christakis D. Characteristics of the informal curriculum and trainees' ethical choices. *Acad Med.* 1996;71:624-33.
4. Swick HM. Toward a normative definition of medical professionalism. *Acad Med.* 2000;75:612-6.
5. Wear D, Castellani B. The development of professionalism: curriculum matters. *Acad Med.* 2000;75:602-11.
6. Swick HM, Szenas P, Danoff D, Whitcomb ME. Teaching professionalism in undergraduate medical education. *JAMA.* 1999;282:830-32.
7. The Medical School Objectives Writing Group. Learning Objectives for Medical Student Education—Guidelines for Medical Schools: Report 1 of the Medical School Objectives Project. *Acad Med.* 1999;74:13-8.
8. Murray E, Gruppen L, Catton P, Hays R, Woolliscroft JO. The accountability of clinical education: its definition and assessment. *Med Educ.* 2000;34:871-9.
9. Arnold E, Blank L, Race K, Chaperone N. Can professionalism be measured? The development of a scale for use in the medical environment. *Acad Med.* 1998;73:1119-21.
10. Pennebaker JW, Francis M. Cognitive emotional and language processes in disclosure. *Cognition and Emotion.* 1996;10:601-26.
11. Hodges B, Turnbull J, Cohen R, Bienenstock A, Norman G. Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. *Med Educ.* 1996;30:38-43.
12. Donnelly MB, Sloan D, Plymale M, Schwartz R. Assessment of residents' interpersonal skills by faculty proctors and standardized patients: a psychometric analysis. *Acad Med.* 2000;75:(10 suppl) S93-S95.
13. Ginsberg S, Regehr G, Hatala R, et al. Context, conflict, and resolution: a new conceptual framework for evaluating professionalism. *Acad Med.* 2000;75:(10 suppl)S6-S11.