Maximizing Participation in Peer Assessment of Professionalism: The Students Speak
Carolyn K. Shue, Louise Arnold, and David T. Stern

Abstract

Background
Medical students have unique information about peers’ professionalism but are reluctant to share it through peer assessment.

Method
Students (231 of 375; 62%) in one school replied to a survey about whether various characteristics of peer assessment (e.g., who receives the assessment, its anonymity, implications for the assessment methods being investigated, peer assessment appears to be one of the most promising since peers observe each others’ behavior through frequent, close contact in a variety of settings where faculty may not be present.2–8

Although several peer-assessment systems have been successfully implemented for medical students and resident physicians,3,9–11 learners have refused to assess each other or have expressed reluctance about participating in peer assessment.12–16 Their reticence may have diminished the reliability of their assessments.14,17 The sociocultural climate surrounding peer assessment may compromise the validity of the assessments in the eyes of students.13 Thus, understanding students’ views about the characteristics of peer assessment that would promote or deter their honest participation in assessing each other is warranted if medical educators are going to be able to involve students in the assessment of professionalism and thereby utilize a valuable assessment resource.

Accordingly, the authors conducted focus groups with a small number of students to identify those characteristics.18 From the students’ perspective, their honest participation in peer assessment is not just a matter of administering an instrument they find acceptable. Rather, their participation depends upon the institutional context in which the instrument is used and the characteristics of the assessment system itself: who does the assessment and who receives it, its purpose, its implications for the peer assessed and for the student evaluator, whether it is required or optional, whether it is anonymous or signed, its timing, and the setting in which it occurs. The purpose of this study is to ascertain if these characteristics of peer assessment generalize to a larger number and greater variety of students.

This study, then, sought to answer the following research questions:

1. Are students willing to participate in peer assessment of professionalism?
2. What characteristics of peer assessment make it easier for students to tell other people about peers’ professional and unprofessional behavior? What characteristics prevent students from reporting their observations?
3. Are there differences among year levels regarding students’ preferences for characteristics of peer assessment? Are there differences between male and female students regarding their preferences?

Method
This research, with IRB approval, was conducted at a midwestern, state-supported medical school with a six-year BA/MD degree curriculum. The school uses peer assessment in a pharmacology course and in obstetrics–gynecology and internal medicine clerkships. It also has an active Honor Council. Although peer assessment is part of the school’s curriculum, students are dissatisfied with the current system. For example, they are concerned that students do not honestly assess each other and use peer assessment as a vendetta. Thereby the full potential of peer assessment has not been realized.

The authors constructed survey items based on emergent themes from student focus groups.18 Fifty-eight survey items asked students to indicate, using a Likert scale, whether various characteristics would prevent (scale value of 1) or encourage (scale value of 5) their honest
participation in a peer assessment system when evaluating unprofessional and then professional behavior. Seven specific characteristics were explored: who receives the assessment (12 total survey items), timing (six items), anonymity (six items), setting (eight items), optional/required participation (four items), and implications for the classmate (10 items) and the reporting student (two items). Other characteristics (12 items) were explored for their potential to facilitate or deter students’ honest participation in peer assessment regardless of whether the assessment focused on professional or unprofessional behavior. A final item, using a three-point ordinal scale, elicited students’ overall willingness to assess each other’s professionalism.

Procedures
An external consultant, with expertise in survey research and quantitative methods, reviewed the psychometric design of the survey. The survey was revised, piloted on a small group of students, and revised further prior to administration. Students were offered the opportunity to be entered into a drawing for one of two $50 gift certificates for each year level. Surveys were administered at the end of the academic year electronically via WebCT or by paper during class meetings. Students provided their consent, but no names were linked to data collected. For electronic submissions, student names were collected in a database separate from survey responses without a means of linking names to responses. Those students answering paper surveys turned in their completed consent form separate from their completed survey directly to the researchers. Students who did not complete paper surveys were asked to complete the survey electronically. If the students chose not to, there was no further follow-up.

Analyses
Kruskal-Wallis and Mann-Whitney nonparametric tests were used to ascertain year and gender differences in students’ responses to the item eliciting their overall views of peer assessment, measured on a three-point ordinal scale. Descriptive statistics, specifically means and standard deviations, were used to explore student preferences for characteristics of peer assessment, measured on a five-point Likert scale. Analysis of variance and Tukey post hoc tests were used to examine year level differences. Independent t tests were used to examine gender differences. The p value for these tests was set at .01 to minimize the potential for spurious results.

Results
Preclinical, transitional, and clinical students (231 of 375; 62% response rate) responded to the survey. At the time of the survey, the Preclinical 1 students (n = 63; 72% response rate) had completed courses for their baccalaureate degree, several medical school basic science requirements, and introductory courses in medicine. The Preclinical 2 students (n = 73; 76%) had finished most of their medical school basic science requirements, a physical diagnosis course, and a full year of continuing care clinic. The Transitional students (n = 52; 58%) had completed basic science, baccalaureate, and a year of continuing care clinic work and family and internal medicine rotations. The Clinical students (n = 35; 34%) had finished another internal medicine rotation and clerkships in obstetrics–gynecology, pediatrics, psychiatry, and surgery. The numbers of female (146) and male (79) participants reflected the gender composition of the school. (Six students did not report their gender.)

Reliability analysis of the survey items resulted in a Cronbach’s alpha of .836 when students were asked their preferences for reporting unprofessional behavior, .906 on items targeting the assessment of professional behavior, and .785 on items measuring other characteristics regardless of whether unprofessional or professional behavior is assessed.

Research Question 1
Sixty-six percent of the students agreed there should be peer assessment of professionalism as long as it reflected their preferences for characteristics of the assessment. Another 22% of the students understood the need for peer assessment, yet felt uncomfortable participating; while only 12% of students felt peers should not be held responsible for the assessment of other students’ professional behavior. Nonparametric tests indicated no significant year level or gender differences regarding students’ willingness to assess their peers’ professionalism.

Research Question 2
Table 1 identifies the characteristics that would most prevent and most encourage honest participation in peer assessment. Mean ratings indicated that students generally rated all characteristics favorably when faced with assessing a peer who acted professionally (range, 3.17–4.33). However, when assessing a peer who had acted unprofessionally, students’ ratings showed greater spread (range, 1.70–4.61). Despite the increased variability in ratings of characteristics for assessing unprofessional behavior, students identified the same characteristics of timing (i.e., end of year), anonymity (i.e., name on evaluation), and setting (i.e., behavior outside of classes and rotations) that would discourage honest participation in the assessment of both unprofessional and professional behaviors. Students identified another set of characteristics of timing (i.e., immediately after the classmate behaves unprofessionally or professionally), anonymity (i.e., 100% anonymous), and setting (i.e., in all clinical rotations) that would promote the honest assessment of unprofessional and professional behaviors. Implications of the assessment for the student evaluator were important to students; whether they assessed a peer’s unprofessional or professional behavior, they strongly preferred that the assessment would not disrupt the relationship they had with the peer assessed.

Student preferences for who receives the assessment and implications for the classmate varied depending on whether they assessed unprofessional or professional behavior (Table 1). When they assessed unprofessional peer behavior, students most preferred that an impartial counselor receives their assessment and that the peer receives instruction to correct the behavior. But when they assessed professional behavior, students most preferred that the peer receives their feedback directly and that their information is used to select members of a professionalism honor society. Regarding characteristics they least preferred, when students assessed unprofessional behaviors, they rated telling the classmate directly and the evaluation affecting the classmate’s grade the lowest; and when they assessed
professional behaviors, they rated an impartial or student advocate as recipient of the assessment and no follow-up or implications for the peer the lowest. Students’ preferences for whether the assessment should be required or optional were unclear.

Additionally, other characteristics of peer assessment that might make it more acceptable to students regardless of whether they were rating peers’ unprofessional or professional behavior were studied. Ratings showed that peer assessments that ask students to evaluate both unprofessional and professional behavior of peers (mean = 4.02, SD = 1.01) or only professional peer behaviors (mean = 3.48, SD = 1.12) would be more likely to promote honest participation than assessments that focus only on peers’ unprofessional behaviors (mean = 2.65, SD = 1.24). Students preferred to assess the behaviors of peers in their own class (mean = 3.76, SD = 1.16) rather than having older students evaluate the behavior of younger students (mean = 3.22, SD = 1.27).

Students also shared their views about the context for the assessment—those institutional factors that could prevent or encourage their participation. Items addressing institutional support were among the most highly rated characteristics. Students indicated that clearly defined professional behavior expectations from faculty members (mean = 4.33, SD = .89), appropriate institutional action based on information provided in peer assessments (mean = 4.33, SD = 1.03), processes by which faculty and residents evaluate their peers (mean = 4.28, SD = .89), and training regarding the giving and receiving of feedback (mean = 4.04, SD = 1.05) were characteristics that would promote student participation in peer assessment.

Research Question 3
Analysis of variance showed one difference among year levels regarding students’ characteristic preferences. Transitional students were more likely to see the characteristic the classmate receives the report of unprofessional behavior directly from the peer (F3,198 = 5.419, p = .001) as preventing honest peer assessment. The Tukey post hoc test (p = .001) indicated that Transitional students (mean = 2.07, SD = 1.37) significantly differed from the Preclinical 1 students (mean = 3.19, SD = 1.56). For all other characteristics, there were no preference differences among the year levels at p < .01.

Independent t tests resulted in statistically significant differences between male and

---

### Table 1
**Students’ Least and Most Preferred Characteristics for Peer Assessment of Professional Behavior**

<table>
<thead>
<tr>
<th>Characteristic categories</th>
<th>Assessment of unprofessional behaviors</th>
<th>Assessment of professional behaviors</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Who receives the assessment</strong></td>
<td>The classmate him/herself</td>
<td>Impartial or student advocate</td>
<td>2.75 (1.44)</td>
<td>4.05 (1.07)</td>
</tr>
<tr>
<td><strong>Timing</strong></td>
<td>At the end of the year</td>
<td>At the end of the year</td>
<td>2.55 (1.29)</td>
<td>3.46 (1.38)</td>
</tr>
<tr>
<td><strong>Counselor or student advocate</strong></td>
<td>3.67 (1.13)</td>
<td>The course instructor, director, or attending</td>
<td>4.32 (1.95)</td>
<td>4.31 (1.98)</td>
</tr>
<tr>
<td><strong>Behavior of peers</strong></td>
<td>Immediately after the classmate behaved unprofessionally</td>
<td>Immediately after the classmate behaved professionally</td>
<td>3.90 (1.22)</td>
<td>4.31 (1.98)</td>
</tr>
<tr>
<td><strong>Anonymity</strong></td>
<td>Name was on the evaluation classmate received</td>
<td>Name was on the evaluation classmate received</td>
<td>1.70 (1.13)</td>
<td>3.17 (1.50)</td>
</tr>
<tr>
<td><strong>Required/optional</strong></td>
<td>Required</td>
<td>Required</td>
<td>3.38 (1.38)</td>
<td>3.51 (1.36)</td>
</tr>
<tr>
<td><strong>Setting</strong></td>
<td>Evaluation included behavior outside of classes and rotations</td>
<td>Evaluation included behavior outside of classes and rotations</td>
<td>2.70 (1.40)</td>
<td>3.20 (1.42)</td>
</tr>
<tr>
<td><strong>Implications for classmate</strong></td>
<td>A component of all clinical rotations</td>
<td>A component of all clinical rotations</td>
<td>3.72 (1.11)</td>
<td>3.87 (1.08)</td>
</tr>
<tr>
<td><strong>Implications for me</strong></td>
<td>Evaluation affects classmate’s grade</td>
<td>No follow-up or implications</td>
<td>2.40 (1.32)</td>
<td>3.22 (1.37)</td>
</tr>
<tr>
<td></td>
<td>Classmate receives corrective instruction</td>
<td>Evaluation helps select members of a professionalism honor society</td>
<td>3.94 (1.01)</td>
<td>3.97 (1.21)</td>
</tr>
</tbody>
</table>

**Scale:** 1 — Prevent, 2 — Somewhat Prevent, 3 — No Impact, 4 — Somewhat Encourage, 5 — Encourage. A complete copy of the survey can be viewed at <http://www.umkc.edu/medicine/omer/projects/index.htm>.

*For this category, students rated one item; thus the high rating indicates a high preference only, not a most preferred choice among a variety of “Implications for me.”*
female students’ preferences for characteristics that targeted interpersonal relationship issues. For reporting professional behavior, women rated the characteristics sharing the information directly with the classmate \(t = 2.83, p = .005\); mean for women = 4.36, SD = 1.13; mean for men = 3.85, SD = 1.32) and the desire that the feedback not disrupt their current relationship with the student \(t = 3.01, p = .003\); mean for women = 4.46, SD = .90; mean for men = 4.00, SD = 1.08) higher than did men. For reporting unprofessional behavior, women rated the characteristic providing their name on the assessment lower than did men \(t = 3.31, p = .001\); mean for women = 1.50, SD = .96; mean for men = 2.08, SD = 1.33).

**Discussion**

Survey results indicate that students support peer assessment and prompt some provisional recommendations for promoting students’ assessment of each other’s professionalism. First, regardless of whether they evaluate a peer behaving unprofessionally or professionally, students prefer that the system has these characteristics: (1) the peer receives the assessment immediately after his/her behavior occurred; (2) the assessment is 100% anonymous; (3) the assessment is part of all clinical rotations; and (4) reporting the classmate’s behavior would not disrupt the student evaluator’s relationship with the peer. Who receives the assessment, its anonymity, and implications of the assessment for the classmate and the evaluator are critical, according to student ratings. But, if an institution were to adopt student preferences for who receives the assessment and its implications for the classmate, the system would have different characteristics dependent upon whether unprofessional or professional peer behavior was assessed. Further, the matter of whether peer assessment should be required remains unresolved without clear guidance from the students who participated in this survey.

Second, students across the year levels generally agreed about the most highly rated characteristics that should be part of a system (see Table 1). Thus, it may be possible to design a system that applies to students at all year levels.

Third, female students may be especially reticent about participating in systems that negatively impact other students or themselves. Further research is needed to fully understand potential gender differences.

Fourth, students believe that institutional support for peer assessment is crucial to the success of the system. Such support includes clear definition of the professional behaviors faculty members expect of students, training for giving and receiving feedback, appropriate institutional action when students provide information about peers’ behavior, and participation of faculty and residents in the assessment of their colleagues’ professionalism. A system of peer assessment that asks students to evaluate both unprofessional and professional behaviors of peers or only professional peer behavior is more likely to promote honest participation than one that asks students to evaluate only unprofessional peer behavior.

These recommendations are provisional due to the following limitations of this work. First, the response rates of the students in the transitional and clinical years were modest. Greater representation of the student body would increase the confidence in generalizing the results to all students in the school. However, the survey results generally coincide with results derived from focus groups. Second, the characteristics of a system encouraging honest assessment of peers’ behavior are statistically derived (based on mean ratings) and should be verified by students themselves. Third, as with all survey research, the responses students gave may reflect their notions of social desirability or their need to criticize the prevailing circumstances surrounding professional behavior and its assessment. A next step to address that limitation could entail a revised peer assessment system incorporating student recommendations derived from this work along with an examination of students’ attitudes and behaviors before and after the new system was implemented. Finally, the structure of the school that these student participants attend may contribute to the general absence of year differences. The structure fosters interaction among students across the years through membership on a longitudinal health care team composed of students from four curricular years. Whether these results generalize to other schools, especially four-year schools that have not yet implemented peer assessment, remains an important but unanswered question that the authors are exploring in a survey of seven four-year institutions.

**Conclusion**

This research provides support for the further development of peer assessment of medical students’ professionalism and highlights the sensitivity of this kind of assessment to contextual factors. As a group, the study participants are willing to engage in peer assessment of professionalism provided the appropriate institutional support, anonymity, faculty oversight, timely assessment, counseling or commendation of peers, and protection for the student evaluator are present. Future research on the practical implementation of these recommendations is warranted.

This project was funded in part by a National Board of Medical Examiners (NBME®) Edward J. Stemmler, MD Medical Education Research fund grant. The project does not necessarily reflect NBME policy, and NBME support provides no official endorsement.

**References**


