A Novel Approach to Assessing Professionalism in Preclinical Medical Students Using Multisource Feedback Through Paired Self- and Peer Evaluations

Amanda R. Emke, Steven Cheng, Ling Chen, Dajun Tian & Carolyn Dufault

Published online: 12 May 2017.

Submit your article to this journal ⍈

Article views: 70

View related articles ⍈

View Crossmark data ⍈

Routledge
Taylor & Francis Group

GROUNDWORK

Check for updates

A Novel Approach to Assessing Professionalism in Preclinical Medical Students Using Multisource Feedback Through Paired Self- and Peer Evaluations

Amanda R. Emke[a], Steven Cheng[b], Ling Chen[c], Dajun Tian[c], and Carolyn Dufault[d]

[a]Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri, USA; [b]Department of Internal Medicine, Division of Renal Diseases, Washington University School of Medicine, St. Louis, Missouri, USA; [c]Division of Biostatistics, Washington University School of Medicine, St. Louis Missouri, USA; [d]Department of Internal Medicine, Division of Medical Education, Washington University School of Medicine, St. Louis, Missouri, USA

ABSTRACT

Phenomenon: Professionalism is integral to the role of the physician. Most professionalism assessments in medical training are delayed until clinical rotations where multisource feedback is available. This leaves a gap in student assessment portfolios and potentially delays professional development. *Approach:* A total of 246 second-year medical students (2013–2015) completed self- and peer assessments of professional behaviors in 2 courses following a series of Team-Based Learning exercises. Correlation and regression analyses were used to examine the alignment or misalignment in the relationship between the 2 types of assessments. Four subgroups were formed based on observed patterns of initial self- and peer assessment alignment or misalignment, and subgroup membership stability over time was assessed. A missing data analysis examined differences between average peer assessment scores as a function of selective nonparticipation. *Findings:* Spearman correlation demonstrated moderate to strong correlation between self-assessments completed alone (no simultaneous peer assessment) and self-assessments completed at the time of peer assessments ($\rho = .59$, $p < .0001$) but weak correlation between the two self-assessments and peer assessments (alone: $\rho = .13$, $p < .013$; at time of peer: $\rho = .21$, $p < .0001$). Generalized estimating equation models revealed that self-assessments done alone ($p < .0001$) were a significant predictor of self-assessments done at the time of peer. Course was also a significant predictor ($p = .01$) of self-assessment scores done at the time of peer. Peer assessment score was not a significant predictor. Bhapkar's test revealed subgroup membership based on the relationship between self- and peer ratings was relatively stable across Time 1 and Time 2 assessments ($\chi^2 = 0.83$, $p = .84$) for all but one subgroup; members of the subgroup with initially high self-assessment and low peer assessment were significantly more likely to move to a new classification at the second measurement. A missing data analysis revealed that students who completed all self-assessments had significantly higher average peer assessment ratings compared to students who completed one or no self-assessments with a difference of –0.32, 95% confidence interval [–0.48, –0.15]. *Insights:* Multiple measurements of simultaneous self- and peer assessment identified a subgroup of students who consistently rated themselves higher on professionalism attributes relative to the low ratings given by their peers. This subgroup of preclinical students, along with those who elected to not complete self-assessments, may be at risk for professionalism concerns. Use of this multisource feedback tool to measure perceptual stability of professionalism behaviors is a new approach that may assist with early identification of at-risk students during preclinical years.

Introduction

Professional behavior is an essential trait in the practice of medicine. It is a cornerstone of not only the physician–patient relationship but the relationship between colleagues working together in the multidisciplinary care of patients. Because of this, professionalism has become an important component of medical education at all stages of training. Nearly 90% of medical schools have reported the integration of professionalism into their curricula, with the majority offering activities ranging from white-coat ceremonies to lectures during the early preclinical years of undergraduate medical education.[1]

Although instruction in professionalism has been widely adopted, assessment of professionalism has been more difficult to implement.[2] Faculty assessors have traditionally been the assessors of students' professionalism

behaviors. However, those in supervisory roles, particularly faculty leading large classroom-based experiences, may lack meaningful opportunities to directly observe individual students' or trainee's professional behaviors.[3] Expanding the range of assessors to a wider population of individuals who interact with students and trainees provides a more complete picture of professional behaviors.[4] This concept, referred to as multisource feedback (MSF) has a long history of development and implementation across a wide array of contexts, including the practice of medicine. Specifically, this method of assessment has been validated as a tool to identify those who may have deficiencies in professional behavior and may respond to early correction.[5–7] Early detection of those at risk for professionalism concerns is a necessary first step to providing early, targeted remediation. The question of the most appropriate assessors, though, remains pivotal for validity and reliability. In the clinical years of medical training, students interact over time with the same nurses, residents, faculty, and peers, allowing multiple sources of feedback on professional behaviors. In the preclinical years of training, students are limited to longitudinal interactions with faculty, peers, and in some cases teaching assistants. Of these interactions, peers have the greatest degree of exposure to one another's professional behaviors over time, and peer assessment has been identified as valid and reliable in many areas of medical training.[8–16] Self-assessment has met with mixed results, with some literature recommending it as a tool for self-directed learning but many studies highlighting the limitations of self-assessment in identifying underperformance, especially by those with the worst performance.[8,9,14,17–21]

This study examined the use of formative self- and peer assessments, with a specific focus on the alignment of those assessments, as a novel multisource professionalism feedback tool. These multisource assessments evaluated professional behaviors observed by peers and self during Team-Based Learning (TBL) activities, which occurred during multiple time points within the required 2nd year of a traditional medical degree curriculum. Studies have suggested that assessments of professionalism should be rooted in an appropriate context for observation of professional behaviors.[22–24] As such, TBL activities were chosen because of the unique context in which evaluation and refinement of professional conduct can be naturally integrated into the collaborative learning experience. While working together to learn material, apply knowledge, and problem solve across a variety of clinical topics, students are required to demonstrate teamwork, participation, active listening, feedback, and discussion—all aspects of professional conduct.[25] In this study, TBL provided an ideal opportunity to allow for context-specific evaluation of professionalism behaviors by those individuals closest to the action: the students and one's self. Perfect alignment between how one self-evaluates and how one's peers evaluate a given situation likely indicates an accurate and useful perception of one's own professional attributes. On the other hand, misalignment—particularly over multiple evaluation instances—likely indicates a consequential perceptual error. Without early identification and remediation, those students who do not realize their own professionalism lapses are most likely to perpetuate them into their future training and work. This study seeks to demonstrate that perceptual errors related to professionalism behaviors can be detected early in medical training through repeated instances of context-specific multisource feedback, creating the conditions necessary for early, targeted remediation.

Methods

Self- and peer assessment instruments

This study was reviewed by the Washington University Institutional Review Board and granted exempt status. All student educational data were received in a fully deidentified form prior to analysis by the research team.

The primary research data were self- and peer ratings of observed professionalism attributes collected following participation in a series of TBL sessions in two required 2nd-year medical school courses (Pediatrics; Renal and Genitourinary Diseases). The formative assessment instruments, adapted from the Koles Method[26] and provided in Table 1, were composed of nine parallel items based on the core attributes of team-based professional behaviors. Students voluntarily and anonymously responded to all items on the self- and peer assessments using a 9-point scale. The scale ranged from 1 (a behavior was never demonstrated) to 9 (a behavior was always demonstrated). Self-assessment data were collected twice per course—first after an initial TBL session *without* simultaneous peer assessment, and again following a final TBL session *with* simultaneous peer assessment. Thus, in a given year across two courses, each student had opportunities to self-assess professionalism behaviors four times and provide peer assessments twice. Completion of and scores on assessments had no impact on student grades.

TBL sessions

The format and objectives of TBL activities, two per course, were presented to the students in both courses through syllabus materials and introductory lectures.

Table 1. Self- and peer assessment questions.

	Peer Assessment	Self-Assessment
1	[Student name] is actively involved in TBL discussions.	I was actively involved in my team's discussions.
2	[Student name] actively shares understanding and knowledge during team discussions.	I actively shared my personal understanding and knowledge during team discussions.
3	[Student name] asked useful or probing questions during team discussions.	I asked useful or probing questions during team discussions.
4	[Student name] had a good balance of listening and participating in team discussions.	I had a good balance of active listening and participation in team discussions.
5	[Student name] demonstrates understanding of the material.	I was more confident in my understanding of the material after today's session.
6	[Student name] is well prepared.	I was well prepared for today's session.
7	[Student name] identifies limitations in understanding.	I identified gaps in my knowledge or areas for improved personal understanding after today's session.
8	[Student name] provides instructive feedback to teammates.	I gave instructive feedback to my teammates during team discussions.
9	[Student name] accepts instructive feedback from teammates.	I accepted instructive feedback from my teammates during team discussions.

TBLs comprised three components: a closed-book individual readiness assurance test, a closed-book group readiness assurance test, and an open-book application exercise. Students were given preparatory materials and clear session objectives for each TBL at least 1 week prior to the actual exercise. The demonstration of professional behavior and collaboration was clearly listed as an objective for each of the TBL sessions in both courses. Students were randomized into groups of seven and provided with their team number and roster at the start of each course. All teams assembled in a single, large auditorium. Placards with team numbers and rosters were placed on desks that were evenly spaced throughout the auditorium, indicating where each group would assemble. Mobile chairs could easily be arranged around the desk to accommodate all members of the team. A team packet was provided on each desk. Within the packet were a closed-book individual readiness assurance test, a closed-book group readiness assurance test, GRAT, and application exercises that could be distributed to each member at the appropriate time. Two faculty members were present at each 2-hour TBL activity, where they served as proctors, observers, and discussion moderators.

Study sample

The study sample included all enrolled 2nd-year Washington University School of Medicine medical students ($n = 246$, split evenly between academic years 2013–2014 and 2014–2015). All students completed at least one of the six self- and peer assessments just described.

Outcomes

Three primary research variables were formed from individual self- and peer assessment items for each course: (a) self-assessment alone (SAA), which was the average rating of the nine individual self-assessment items completed when no peer assessment was made; (b) self-assessment (SA), which was the average rating of the nine self-assessment items completed at the same time that peer assessments were made; and (c) average peer assessment (PA), which was the average peer rating on all nine items for a given individual, averaged across all team peers' ratings. TBL teams in both courses were formed using published TBL methods[26] and comprised six or seven students, so the average PA ratings for each student were composed of five or six individual peer ratings per course.

Controlled variables

We included multiple covariates to account for potential effects of mediating variables on the primary dependent variable, average SA rating. These included a year covariate to account for any potential cohort differences between the 2013–2014 and 2014–2015 academic years. One known difference was the order of the courses across both years; in 2013–2014, the Renal and Genitourinary Diseases course preceded the Pediatrics course. The order was reversed in 2014–2015, and the year covariate allowed us to examine the effects of this reversal. Therefore, the temporal order of the two courses was assessed in recognition of the potential effects of assessment exposure, intellectual growth, and other factors on self-assessment. Similarly, course and gender covariates allowed for examination of the effects of these potential mediators.

Analysis

Spearman's rank correlation coefficient was used to characterize the pairwise correlation among SAA, SA, and PA. Considering the data were not normally distributed and repeated measurements were obtained, generalized estimating equation models were used to examine the effects of various predictor variables on the dependent variable, SA. These predictor variables included SAA, PA, course (Renal and Genitourinary Diseases/

Table 2. Self- and peer assessment subgroup assignment rules.

Subgroup	Subgroup Classification Assignment Rule
High Peer/High Self (Accurate Self-Perception)	Individual peer assessment score was greater than the median peer assessment score *and* the individual self-assessment score was greater than the median self-assessment score.
Low Peer/Low Self(Accurate Self-Perception)	Individual peer assessment score was less than the median peer assessment score *and* the individual self-assessment score was less than the median self-assessment score.
High Peer/Low Self (Inaccurate Self-Perception)	Individual peer assessment score was greater than the median peer assessment score *and* the individual self-assessment score was less than the median self-assessment score.
Low Peer/High Self (Inaccurate Self-Perception)	Individual peer assessment score was less than the median peer assessment score *and* the individual self-assessment score was greater than the median self-assessment score.

Note. Only self-assessment made at the time of peer-assessment was used in the determination of subgroup membership.

Pediatrics), year (Academic Year 2013–2014 and 2014–2015) and stage (Time 1 and Time 2).

To examine the stability of self- and peer assessments over multiple time points, individuals were classified into one of four subgroups based on their deviation from the self- and peer assessment medians at each stage. The classification rules underlying the formation of each group are listed in Table 2. Bhapkar's goodness of fit test is an extension of McNemar's test and used to test goodness of fit in situations when each group has more than two levels. With four levels in each stage for this study, Bhapkar's test[27] was used to determine whether overall grouping classification at Time 1 differed from grouping classification at Time 2.

A missing data subgroup analysis was also conducted. Students were placed into one of four groups based on having no missing self-assessments, missing both assessments, or missing only one assessment (SAA or SA). Average peer assessment ratings were compared among these groups by general linear model.

All statistical analyses were conducted using SAS 9.3,[28] two-sided with a significance level of 0.05.

Results

Table 3 provides the basic descriptive statistics for the full study sample organized by year, course, and time for average SAA, SA, and PA scores. Although the full range of the rating scale was used for all three variables, ratings

Table 3. Mean self- and peer assessment scores by cohort, stage, and course.

	M Assessment Score (SD)		
	Peer Assessment	Self-Assessment at Time of Peer Assessment	Self-Assessment Alone
Cohort 1 (2013–2014)			
Time 1: Renal	8.0 (7.6)	7.9 (1.2)	7.5 (0.9)
Time 2: Pediatrics	8.5 (0.6)	8.6 (0.7)	8.0 (0.9)
Cohort 2 (2014–2015)			
Time 1: Pediatrics	8.5 (0.5)	8.5 (0.6)	8.0 (0.7)
Time 2: Renal	8.3 (0.3)	8.2 (0.7)	7.6 (0.7)

at the highest end of the scale were the most common. Given this non-normal distribution of data, Spearman's rank correlation was used to examine the relationships between SAA, SA, and PA. All three variables were statistically significantly correlated with each other. The strongest correlation was between SAA and SA ($\rho = .59$, $p < .0001$), whereas the relationships between SAA and PA ($\rho = .13$, $p = .01$) and SA and PA ($\rho = .21$, $p < .0001$) were weaker in comparison.

Predicting self- and peer assessment scores

Generalized estimating equation was used to examine variables that were predictors of the SA rating, and the relative change in SA was examined. SAA ($p < .0001$) was a significant predictor, revealing that for every 1-point increase in SAA, the SA rating increases by 7.5%, 95% confidence interval (CI) [5.7%, 9.3%]. Course was also a significant predictor ($p = .01$). SA in the Pediatrics course was 2.5%, 95% CI [0.5%, 4.4%] higher than in the Genitourinary Diseases course. Notably, PA was not a significant predictor of SA, suggesting that there are important discrepancies between how students rate themselves and how their peers rate the same individuals. There were no significant effects of the year, gender, or stage variables in the model in univariate or multivariate analysis.

Subgroup classification stability

Table 4 presents the frequency and stability of individual classifications into subgroups based on the relationship between PA and SA scores. Bhapkar's test revealed that in the overall model, classification of individuals was relatively stable across Time 1 and Time 2 assessments ($\chi^2 = 0.83$, $p = .84$). This overall stability was evident in three of the four subgroups but did not characterize individuals in the low-PA/high-SA subgroup at Time 1 (Group 4). Unlike their counterparts in the other Time 1 subgroups, Group 4 members were numerically most

Table 4. Subgroup frequency and stability.

	Time 2				
	High Peer/ High Self	Low Peer/ Low Self	High Peer/ Low Self	Low Peer/ High Self	Total
Time 1					
High Peer/ High Self	20	5	6	7	38
Low Peer/ Low Self	9	14	6	6	35
High Peer/ Low Self	2	8	14	4	28
Low Peer/ High Self	7	13	2	6	28
Total	38	40	28	23	129

likely to be reclassified into a new subgroup (low-PA/low-SA) at Time 2.

Self-assessment selective noncompliance

As seen in Table 3, all students in the sample received peer assessments. However, some students had incomplete self-assessments, missing alone, at the time of peer, or all self-assessments. We predicted that students who opted to not complete one or both self-assessments would have lower average PA ratings, and we confirmed this hypothesis. As shown in Table 5, individuals who skipped both self-assessments had significantly lower PA scores compared to those who completed both of them, with a difference of –0.32, 95% CI [–0.48, –0.15].

Discussion

Assessment during medical training is important for benchmarking performance, identifying strengths and weaknesses, and driving learning.[29–33] Competencies such as teamwork, professionalism, communication, and interpersonal skills are less likely to be directly observed by a supervisor but can be observed and assessed by other people with whom trainees interact.[32,34] As such, MSF has been identified as the optimal mechanism for assessing these behavioral competencies.[32–36] This study used MSF assessments of professional behaviors exhibited by students during multiple, required TBL exercises.

Table 5. Mean peer assessment scores for individuals by compliance status for completing self-assessments.

Self-Assessments Completed	N	M Peer Assessment (SD)
Both Self-Assessments	275	8.4 (0.6)
Self-Assessment Alone Only	80	8.3 (0.7)
Self-Assessment at Time of Peer Only	71	8.3 (0.6)
No Self-Assessments Completed	63	8.1 (0.8)

The purpose of this study was to characterize the patterns and stability of the relationship between how students assessed their own behaviors, relative to how they were perceived by their peers. By using MSF to identify alignment in the perceptions of professional behavior, this study demonstrated that it is possible to consistently identify students who do not have accurate perceptions regarding their behaviors, making it likely that they will neither seek nor receive targeted remediation in this core professional competency during their preclinical years.

The role of assessment in identification of deficiencies is critical for remediation and performance improvement of both cognitive and affective competencies. Most of the literature discussing early identification of poor performance focuses on knowledge deficits, especially regarding the preclinical years of medical education.[37–41] That said, reports of lapses in professionalism among medical students are not uncommon and range from cheating and plagiarism to misrepresentation of publications, lying about performance, and falsification of documentation.[42] Significant evidence exists detailing the long-standing nature of professionalism problems showing that medical students with lapses in professionalism continue to demonstrate or have persistent deficiencies throughout medical training and practice.[43,44] Of the articles that looked at the need to remediate professional competencies, most addressed clinical students and specifically addressed clinical decision-making or problem-solving competencies.[37,38,41] These findings highlight the importance of professionalism assessment and, more importantly, early assessment for early identification and remediation.

The concept of assessing professionalism in preclinical medical students is not novel. Phelan and colleagues in 1993 and Papadakis and colleagues in 2001 described the development and implementation of preclinical professionalism assessments.[45] Unfortunately, routine implementation of preclinical professionalism assessments, especially MSF for assessment of professionalism, is not widely reported. In fact, preclinical assessment of professionalism, as reported by Ziring and colleagues, focuses on incident-based reports of professionalism lapses with less than 50% of the schools reporting professionalism lapses being identified by formal peer assessment or items on routine student course evaluations.[38]

This study addresses some of the challenges of preclinical professionalism assessment by offering a unique MSF method for early identification of students with at-risk professionalism behaviors through careful selection of the context and assessor. The choice of self- and peer assessment within small-group instructional methods has been described in the Problem-Based Learning

setting, an instructional method similar to TBL in the use of clinical cases to promote collaborative learning and problem-solving skills. Those reports used MSF in the form of self-, peer, and tutor assessments to assess affective competencies in preclinical medical students. Although this body of work indicates some variability in outcomes, students were shown to underrate themselves, whereas peers overrate compared to tutors, which were used as the "gold standard."[8,14]

The format of TBL, however, does not facilitate direct interaction between a tutor and individual students. Despite the lack of a tutor as a gold standard, an important and broadly applicable contribution of this work is the thoughtful identification of valid candidates for providing assessment because direct observation is an integral component of accurate assessment of affective competencies.[5,7,17,24] This study accurately identifies the assessor with the ability to directly observe professional behaviors in preclinical students—peers and the student.

The prolonged interactions with peers and ability to create a composite or average score, as done in this study, has previously demonstrated moderate to high internal consistency and interrater reliability.[8–16] Peer assessments have also been shown to correspond to later evaluations by faculty and program directors, making them an important component of MSF for early identification of students needing remediation.[11,13] Despite these strengths, many studies elucidate challenges with peer assessment. First, peer assessments often demonstrate a halo effect because students select the peers tasked with completing assessments.[9,13] Second, peer assessments demonstrate a ceiling effect or clumping of scores, more so when scores are used in a summative manner, thus limiting their usefulness by preventing the identification of unprofessional students.[8,12,13,46] Regardless of the benefits of peer assessment, these limitations again highlight the importance of multiple assessors for valid and reliable early identification of at-risk students. The addition of self-assessments to peer assessments of professional behaviors in this study creates multiple sources of observation, a necessity for the creation of MSF.

Self-assessment, considered an integral component of MSF and professional development,[47] does not reach its full potential, though, when used in isolation. Medical student accuracy with self-assessment compared to peer or faculty ratings is low when looking at interview skills, identification of strengths and weaknesses, group participation, and problem solving.[17] The reports from PBL, as stated earlier, demonstrated mixed reviews of self-assessment compared to other assessments.[8,14] Several studies comparing surgical residents' self-assessments to the assessments of others (nurses, peers, and faculty) have demonstrated inflation of self-ratings in a variety of domains.[18–20] Even more concerning are the studies demonstrating that the most incompetent individuals have the worst self-assessment skills and the increased disadvantage of self-assessing interpersonal skills.[9,17,19,21,48] By using the alignment or misalignment of self- and peer assessments, this study provides a unique method that capitalizes on the inaccuracy of self-assessment. In addition, the focus on perceptual errors of professional behaviors provides a generally applicable method for early detection of at-risk students. The longitudinal and formative nature of the paired assessments used in this study allows students an opportunity to recognize their performance gaps. As previously identified in the literature, this aspect of the work described here allows paired assessments to promote self-directed learning and performance improvement, as well as improved ability to self-assess over time.[17,32,47]

Conclusions

Guided by the conceptual framework of MSF, the use of formative paired professionalism assessments following TBL sessions in this investigation addressed many of the concerns raised by previous studies of professionalism assessment. Despite the limitations of being a single institution with a traditional curriculum, this study focused on the broadly applicable aspects of identifying the most appropriate context, assessor, and outcome measure for valid and reliable MSF of professional behaviors. First, peers had shared experiences and many opportunities to directly observe the specific professionalism behaviors being assessed, increasing ability to make assessments based on direct observations. Further, TBL activities contextualized professionalism, ensuring that the behavioral actions being assessed were realistic and meaningful. In addition, in the present study, rater biases were minimized and controlled by using randomly assigned TBL teams rather than allowing for student-selected peer raters, as described in other studies of self- and peer assessments.[5] Finally, pairing self and peer professionalism assessments mitigated the shortcomings of only using self- or peer assessments identified in the previous studies by provided a relational (self vs. peers) assessment of performance. This relational element of assessment, a core tenet of MSF, enabled identification of students who were unaware of their unprofessional behaviors. As such, these paired assessments overcame cited limitations of professionalism assessments by incorporating them into a larger context and proved useful in identifying professionally at-risk students.

In addition, this study provides unique mechanisms for the early identification of professionally at-risk students. Previous studies of correlation between self- and peer assessments focus on using the outcomes to address

the validity or accuracy of one of the assessments. In this study, the degree of alignment is, in and of itself, an assessment that identifies students with at-risk professional behaviors. Much of the literature comparing self- to other assessments highlights the frequency with which individuals provide lower self-assessments compared to peers or other assessors.[8,14] Therefore, focusing attention on students with high self-assessment scores but low peer-assessment scores serves as a unique early indicator and a tool to identify students whose peers say they saw reasons for concern while a self-assessor saw none. Finally, the absence of participation in all self-assessment correlated with the lowest peer assessment scores. Therefore, identifying students who opt out of self-assessment serves as an additional mechanism for early professionalism detection.

Future directions

Our study identified three groups of students at risk for professionalism concerns, and each group may have unique characteristics. Students in the consistently high-SA/low-PA group appeared to lack insight or awareness into their performance when they did not improve with time or repeated assessments. The other two groups, those that rank themselves low and those that self-select out of self-assessment, appear to have insight into their poor performance but do not improve over time as indicated by continued low peer assessment scores. Understanding the characteristics behind these different at-risk groups would impact educational strategies for improvement and remediation. Adding the Self-Reflection and Insight Scale (SRIS) to student assessments may be an option to assist with exactly this type of identification. The SRIS can be used to measures a student's engagement in reflection, need for reflection, and insight.[49] Administered judiciously, the SRIS could provide an understanding of the characteristics behind each at-risk group and specific targets for remediation.

Dunning and Kruger suggested that assessment accuracy can be improved in some individuals through repeated exposure to assessment.[21] The students from this study who moved from the high-SA/low-PA group to the low-SA/low-PA group support Dunning and Kruger's claim because their accuracy improved without any intervention or knowledge of their peer scores. As such, one solution would be to increase the frequency of professionalism assessments in the preclinical years. In this study, students were not required to complete the professionalism assessments. Making assessments mandatory would increase the number of raters and evaluations thus increasing the interrater reliability and, possibly, accuracy.[12,21] Educators must carefully weigh

the impact of a change to mandatory completion, though, because of the possibility of losing the ability to identify an important group of at-risk students. Reliability could also be improved if the assessment tool was expanded to other educational activities where professional conduct could be demonstrated and directly observed. The risk of these solutions is increased assessment fatigue with the ultimate result of less discriminating evaluations over time without proof of increased performance.[10]

Finally, to truly prove useful, paired self- and peer assessments must accurately identify students with professionalism concerns early in their academic career. Although the assessment strategy in this study identified at-risk groups of students, further studies must be conducted to determine if these students develop professionalism concerns in the clinical setting. As such, the next phase of our research will be to examine the longitudinal relationship between the groups identified as being at-risk for professionalism concerns during their preclinical years with assessments of their observed professionalism behaviors during clinical training years.

References

1. Swick HM, Szenas P, Danoff D, Whitcomb ME. Teaching professionalism in undergraduate medical education. *Journal of the American Medical Association* 1999;282:830–2.
2. Epstein RM, Hundert EM. Defining and assessing professional competence. *Journal of the American Medical Association* 2002;287:226–35.
3. Pulito AR Donnelly MB, Plymale M, Mentzer RM Jr. What do faculty observe of medical students' clinical performance? *Teaching and Learning in Medicine* 2006;18:99–104.
4. Barrett A, Galvin R, Steinert Y, Scherpbier A, O'Shaughnessy A, Horgan M, et al. A BEME review of the use of workplace-based assessment in identifying and remediating underperformance among postgraduate medical trainees. *Medical Teacher* 2016;18:1188–98.

5. Donnon T, Al Ansary A, Al Alawi S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: A systematic review. *Academic Medicine* 2014;89:511–6.

6. Naidoo S, Lopes S, Patterson F, Mead HM, and MacLeod S. Can colleagues', patients', and supervisors' assessments predict successful completion of postgraduate medical training? *Medical Education*. Epub 2016 November 23.

7. Warm EJ, Schauer D, Revis B, Boex JR. Multisource feedback in the ambulatory setting. *Journal of Graduate Medical Education* 2010;2:269–77.

8. Eva KW. Assessing tutorial-based assessment. *Advances in Health Sciences Education* 2001;6:243–57.

9. Arnold L. Assessing professional behavior: Yesterday, today, and tomorrow. *Academic Medicine* 2002;77:502–15.

10. Bryan RE, Krych AJ, Carmichael SW, Viggiano TR, Pawlina W. Assessing professionalism in early medical education: Experience with peer evaluation and self-evaluation in the gross anatomy course. *Annals, Academy of Medicine, Singapore* 2005;34:486–91.

11. Dannefer EF, Henson LC, Bierer SB, Grady–Weliky TA, Meldrum S, Nofziger AC, et al. Peer assessment of professional competence. *Medical Education* 2005;39:713–22.

12. Cottrell S, Diaz S, Cather A, Shumway J. Assessing medical student professionalism: An analysis of peer assessment. *Medical Education Online* 2006;11:8.

13. Lurie SJ, Lambert DR, Nofziger AC, Epstein RM, Grady-Weliky TA. Relationship between peer assessment during medical school, dean's letter rankings, and ratings by internship directors. *Journal of General Internal Medicine* 2007;22:13–16.

14. Papinczak T, Young L, Groves M, Haynes M. An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Medical Teacher* 2007;29:e122–32.

15. Kamp RJA, Dolmans DHJM, van Berkel HJM, Schmidt HG. Can students adequately evaluate the activities of their peers in PBL? *Medical Teacher* 2011;33:145–50.

16. Speyer R, Pilz W, van der Kruis J, Brunngs JW. Reliability and validity of student peer assessment in medical education: A systematic review. *Medical Teacher* 2011;33:e572–85.

17. Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, et al. The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide No. 10. *Medical Teacher* 2008;30:124–45.

18. Rose JS, Waibel BH, Schenarts PJ. Disparity between resident and faculty surgeons' perceptions of preoperative preparation, intraoperative teaching, and postoperative feedback. *Journal of Surgical Education* 2011;68:459–64.

19. Lipsett PA, Harris I, Downing S. Resident self-other assessor agreement. *Archives of Surgery* 2011;146:901–6.

20. Gow KW. Self-evaluation: How well do surgery residents judge performance on a rotation? *The American Journal of Surgery* 2013;205:557–62.

21. Kruger J, Dunning D. Unskilled and unware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 1999;77:1121–34.

22. Morihara SK, Jackson DS, Chun MB. Making the professional curriculum for undergraduate medical education more relevant. *Medical Teacher* 2013;35:908–14.

23. ten Cate O, Chen HC. The parts, the sum and the whole—Evaluating students in teams. *Medical Teacher* 2016;38:639–41.

24. Rees CE, Knight LV. The trouble with assessing students' professionalism: Theoretical insights from sociocognitive psychology. *Academic Medicine* 2007;82:46–50.

25. Parmelee DX, Hudes P. Team-based learning: A relevant strategy in health professionals' education. *Medical Teacher* 2012;34:411–3.

26. Levine RE. Peer evaluation in team-based learning. In L Michaelsen, D Parmelee, KK McMahon, RE Levine (Eds.), *Team-based learning for health professions education: A guide to using small groups for improved learning*. Sterling: Stylus, 2007.

27. Agresti A. *Categorical data analysis*. New York: Wiley & Sons, 1990.

28. SAS® 9.3. *System Options*. 2nd Ed. Cary, NC: SAS Institute, 2011.

29. Norcini J, Burch V. Workplace-based assessments as an education tool: AMEE Guide No. 31. *Medical Teacher* 2007;29:855–71.

30. McSweeney K. *Assessment for learning: From theory to practice: Final report*. September 2012.

31. Holmboe ES. Faculty and the observation of trainees' clinical skills: Problems and opportunities. *Academic Medicine* 2004;79:16–22.

32. Epstein RM. Assessment in medical education. *The New England Journal of Medicine* 2007;356:387–96.

33. Ferris H, O'Flynn D. Assessment in medical education: What are we trying to achieve? *International Journal of Higher Education* 2015;4:139–44.

34. Lockyer J. Multisource feedback: Can it meet criteria for good assessment? *Journal of Continuing Education in the Health Professions* 2013;33:89–98.

35. Epstein RM, Dannefer EF, Nofziger AC, Hansen JT, Schultz SH, Jospe N, et al. Comprehensive assessment of professional competence: The Rochester Experiment. *Teaching and Learning in Medicine* 2004;16:186–96.

36. Al-Wardy NM. Assessment methods in undergraduate medical education. *QU Medical Journal* 2010;10:203–9.

37. Hauer KE, Ciccone A, Henzel TR, Katsufrakis P, Miller SH, Norcross WA, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Academic Medicine* 2009;84:1822–32.

38. Ziring D, Danoff D, Grosseman S, Langer D, Esposito A, Jan MK, et al. How do medical schools identify and remediate professionalism lapses in medical students? A study of U.S. and Canadian medical schools. *Academic Medicine* 2015;90:913–20.

39. Yates J, James D. Predicting the "strugglers": A case-control study of students at Nottingham University Medical School. *British Medical Journal* 2006;332:1009–13. doi:10.1136/bmj.38730.678310.63

40. Cleland J, Arnold R, Chesser Al. Failing finals is often a surprise for the student but not the teacher: Identifying difficulties and supporting students with academic difficulties. *Medical Teacher* 2005;27:504–8.

41. Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. The remediation challenge: Theoretical and methodological insights from a systematic review. *Medical Education* 2013;47:242–51.

42. Fargen KM, Drolet BC, Philibert I. Unprofessional behaviors among tomorrow's physicians: Review of the literature with a focus on risk factors, temporal trends, and future directions. *Academic Medicine* 2016;91:858–64.

43. Papadakis MA, Hodgson CS, Tererani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Academic Medicine* 2004;79:244–9.

44. Papadakis MA, Teherani A, Branch MA, Knetter TR Rattner SL, Stern DT, et al. Disciplinary action by medical boards and prior behavior in medical school. *The New England Journal of Medicine* 2005;353:2673–82.

45. Papadakis MA, Loeser H, Heady K. Early detection and evaluation of professionalism deficiencies in medical students: One school's approach. *Academic Medicine* 2001;76:1100–6.

46. Spandorfer J, Puklus T, Rose V, Vahedi M, Collins L, Giordano C, et al. Peer assessment among first year medical students in anatomy. *Anatomical Sciences Education* 2014;7:144–52.

47. Antonelli MA. Accuracy of second-year medical students' self-assessment of clinical skills. *Academic Medicine* 1997;72:563–5.

48. Tousignant M, DesMarchais JE. Accuracy of student self-assessment ability compared to their own performance in a problem-based learning medical program: A correlation study. *Advances in Health Sciences Education* 2002;7:19–27.

49. Roberts C, Stark P. Readiness for self-directed change in professional behaviors: Factorial validation of the Self-reflection and Insight Scale. *Medical Education* 2008;42:1054–63.