Original Article

Assessing Professionalism in Early Medical Education: Experience with Peer Evaluation and Self-evaluation in the Gross Anatomy Course

RE Bryan,[1]MD, AJ Krych,[1]MD, SW Carmichael,[1]PhD, TR Viggiano,[2]MD, W Pawlina,[1]MD

Abstract

Introduction: As today's healthcare model moves toward more streamlined and corporate industrialism, it is our responsibility, as doctors, to ensure the integrity of medicine's foundation in professionalism. The erosion of professional values not only creates a climate of animosity, but reverberates negatively to impact the development of students, who model their behaviour after those they most respect. This hazard has spurred an evaluation of medical school curricula, with a new emphasis on professionalism in the philosophy of medical education. Courses such as Gross Anatomy that, in the past, offered "pure content," are now being used to teach and evaluate professionalism. The goal of this study was to determine if peer evaluation and self-evaluation used in conjunction and implemented early in the medical curriculum, can serve as useful tools to assess and provide feedback regarding professional behaviour in first-year medical students. Materials and Methods: From 1999 to 2003, students at Mayo Clinic College of Medicine evaluated themselves and their peers during the Gross and Developmental Anatomy Course. Numerical evaluations and written comments were statistically analysed within established categories of professionalism and correlated with academic performance, gender, and peer rating and self-rating. Results: The majority of written comments pertained to inter-professional respect, responsibility, and excellence. Students who gave higher peer evaluation and self-evaluation scores provided more positive comments, and students performing well in the course provided more positive comments about their peers and themselves than did those struggling academically. Students consistently rated their peers higher than themselves, and male students rated themselves higher than did female students. Conclusions: Implementing peer evaluation and self-evaluation early in the medical curriculum is a valuable exercise in teaching first-year medical students assessment skills when evaluating their behaviour, as well as the behaviour of their colleagues.

Ann Acad Med Singapore 2005;34:486-91

Key words: Anatomy, Peer evaluation, Professionalism, Self-evaluation, Teaching

Introduction

The professional role of physicians implies a commitment to upholding social order by providing strong leadership, good moral judgement, and the ethical practice of medicine.[1-3] However, in the latter part of the 20th century, increasing pressure from political, legal, and market forces has effected significant change in healthcare delivery in nearly every industrialised country.[4] Whereas economic drive has funded exponential scientific growth and advancement, these successes have, unfortunately, been accompanied by a compromise of philanthropic values and a growing negative sentiment among physicians. Additionally, it is well-documented that despite an expected increase in medical students' ethical reasoning skills over the course of their training, most experience a static or negative progression.[5-14] This may be attributed to the fact that medicine is increasingly shrouded by required competencies such as "system-based care", "evidence-based medical practice", "outcome-based standards of care", "population medicine", "quality management", and "information access practice". While such guidelines provide structure to corporate endeavours, an abundance of bureaucracy diminishes the humanness of medical practice which is its professionalism – our contractual obligation to

[1] Department of Anatomy
[2] Department of Internal Medicine
Division of Gastroenterology and Hepatology
Mayo Clinic College of Medicine, USA
Address for Reprints: Dr Wojciech Pawlina, Department of Anatomy, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, MN 55905, USA.
Email: pawlina.wojciech@mayo.edu

society. This fundamental compromise undermines professional esteem, and ultimately reverberates to affect medical students.

As such, it is evident that a well-trained and effective physician must be as competent in professionalism and ethical reasoning as he or she is in academics.[15] To this end, medical schools throughout the United States have actively re-examined their curricula, finding that though medical students model their professional behaviour after those they most respect, traditional methods of instilling professional values, such as observation and enculturation, are no longer sufficient.[16] Because professionalism has not historically been emphasized in medical education,[16] accepted methods for teaching and assessment have not been agreed upon. Therefore, medical educators have begun to implement innovative solutions to achieve these ends.[17-23] Currently, almost 90% of US medical schools provide formal activities designed to teach medical professionalism.[16,24]

Several authors have attempted to deconstruct professionalism's core attributes, to determine which aspects are essential to professionalism education.[25-27] Others have focused on determining "success" in teaching professionalism.[17-22,24] Some schools initiate professionalism training from the very beginning of the medical school curriculum, and thus courses such as Gross Anatomy, which were traditionally thought of as "pure science," are now being utilised to teach professionalism as well.

Organisations such as the General Medical Council (GMC) and the American Board of Internal Medicine's (ABIM) "Project Professionalism"[1] have suggested that an important component of the development of medical students' professionalism is self-assessment. In addition, several educators have suggested that peer evaluation may be a useful adjunct.[28,29] Both formats help students develop the ability to make judgements, a necessary skill for study and professional life. By judging the work or behaviour of others, students gain insight into their own performance, an important element of professional competence.

The goal of this study was to determine if peer evaluation and self-evaluation used in conjunction with and implemented early in the medical curriculum, can serve as useful tools to assess and provide feedback regarding professional behaviour in first-year medical students.

Materials and Methods

From 1999 to 2003, five consecutive classes at the Mayo Clinic College of Medicine (n = 213 students) were asked to evaluate professionalism in themselves and their colleagues during the first-year Gross and Developmental Anatomy Course. Students were randomly divided into dissection groups (3 to 4 students/group), with whom they worked intimately for the duration of the class. Students were then asked to complete an online questionnaire (Appendix 1) at the midpoint and conclusion of the course, rating themselves and their group members on a 5-point Likert scale. Students were also asked to provide written comments on positive and negative interactions regarding professional values. Remarks for each student were compiled and returned to that individual in an anonymous fashion, to provide candid, non-punitive feedback and encourage self-reflection. Results of the peer evaluation and self-evaluation had no impact on the student's final grade.

With identifying information removed, each submission was evaluated and coded as positive or negative. Data were recorded according to predetermined aspects of professionalism: Excellence, Compassion, Confidentiality, Integrity, Inter-professional Respect, Responsibility, Self-policing, Accountability, and Miscellaneous. In addition, each comment was classified according to 3 subcategories: peer-provided positive comments, self-provided positive comments, or negative comments. Positive comments included those praising students' professionalism; negative comments reflected areas of perceived need for improvement. Results were also analysed to assess interclass differences within the 5-year sample.

Descriptive statistics were analysed as either mean (with standard deviation), or as median and interquartile percentiles as appropriate. The associations between the number of positive comments and the numerical rating were estimated using Spearman's rank correlation coefficient.

Linear regression was used to assess associations between the dependent factor of academic performance (final course percentage) and peer evaluation, self-evaluation, and the net number of positive comments received by each student and his/her peers. The estimated impact of the predictors from these models with course performance was reported, along with the Pearson correlation coefficient.

The number of positive comments and the final grade received were compared between grading systems (percentage and pass/fail) and between genders, using a two-sample *t*-test, assuming unequal variances. A similar analysis was done comparing peer evaluation and self-evaluation between genders. The change in positive comments and the numerical rating received were compared at the midpoint and course conclusion using a paired *t*-test. Similarly, students' self-rating was compared to the rating received from their peers.

The alpha-level was set at 0.05 for statistical significance.

Results

From the 213 students polled [100 (46.9%) females, 113

(53.1%) males], 1650 evaluations were submitted (participation = 100% of students): 1234 peer evaluations and 426 self-evaluations. Three thousand one hundred and sixty-five written comments related to professional values (2810 positive comments and 355 negative comments) were recorded and categorised, as described above.

Comments on Professionalism

With respect to positive comments (n = 2810), the mean (± sd) was 11.7 (± 3.1) net positive comments per student, with a median of 12 and interquartile range of 10 to 14. The distribution of positive comments in peer evaluation and self-evaluation is indicated in Figure 1. Students received the most praise from their colleagues in the areas of Inter-professional Respect (40.2%), Excellence (21.6%), and Responsibility (16.9%). In self-evaluation, students commented positively in Self-policing (35.7%), Inter-professional Respect (31.4%), and Accountability (13.1%). The number of peer-provided positive comments showed a positive correlation with peer rating (Spearman, r = 0.52, P <0.001). A similar positive, though weaker correlation, was seen between positive self-provided comments and self-ratings (Spearman, r = 0.12, P = 0.08). A positive correlation was also seen between the number of peer-provided positive comments and self-rating (Spearman, r = 0.16, P = 0.02).

With respect to negative comments (n = 355), the mean was 1.67 negative comments per student. Interestingly, comments predominantly addressed failures in Inter-professional Respect (27.9%), Accountability (26.1%), and Self-policing (22%), the same categories emphasised among self-evaluations.

The only category that received neither positive nor negative comments was Confidentiality. It is possible that first-year medical students do not see a need to comment on confidentiality, despite the fact that breach of confidentiality has been reported among students in the gross anatomy laboratory.[30]

Peer Evaluation versus Self-evaluation

Students received significantly more positive comments from their peers than from themselves (Table 1). Students were also ranked higher by their peers than by themselves, with a mean (± sd) of 4.3 (± 0.5) and 3.6 (± 0.8) respectively, P <0.001.

Academic Performance

Linear regression indicated a slight positive correlation between the final grade and the total number of positive comments received (r = 0.22, P <0.001). Each positive comment correlated with an increase of approximately 0.5 points of the final grade. A similar association was seen between the final grade and the numerical rating scale. Peer rating also correlated with the final grade, (r = 0.26, P <0.001), with an associated 3.3-point increase in the final grade per peer-rating point. Self-rating showed a weaker positive correlation, (r = 0.14, P = 0.04), with each point in self-rating associated with an increase of approximately 1.0 final grade point.

Gender Differences

Males received significantly more positive comments than females on peer evaluations [mean (± sd) of 9.1 (± 2.5) and 8.4 (± 2.0) respectively, P = 0.025], and were rated higher than females on peer-provided numerical rating [mean (± sd) 4.4 (± 0.5) and 4.2 (± 0.5) respectively, P = 0.02]. As indicated in Figure 2, males also rated themselves more highly than did females [mean (± sd) 3.7

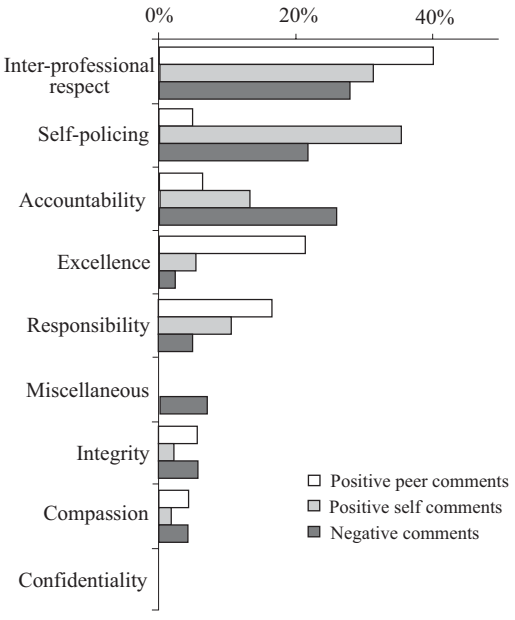

Fig. 1. Distribution of positive and negative comments between predetermined categories of professionalism.

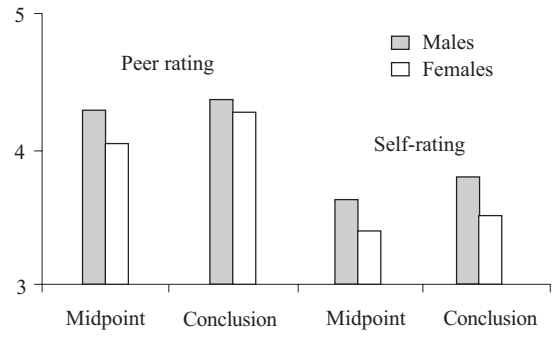

Fig. 2. Average numerical peer and self-rating comparing male versus female scores at the midpoint and conclusion of the Gross and Developmental Anatomy Course.

Table 1. Summary of the Distribution of Written Positive and Negative Comments in Peer Evaluations and Self-evaluations at the Midpoint and Conclusion of the Gross and Developmental Anatomy Course

	Total number and percentage distribution of written comments					
	Positive comments Peer evaluations		Positive comments Self-evaluations		Negative comments Peer and self-evaluations	
	Midpoint	Conclusion	Midpoint	Conclusion	Midpoint	Conclusion
Total number of comments	915	890	550	455	250	105
Excellence	20%	24%	5%	5%	4%	–
Compassion	5%	4%	3%	1%	4%	5%
Confidentiality	–	–	–	–	–	–
Integrity	4%	7%	2%	2%	2%	10%
Inter-professional respect	42%	39%	31%	32%	32%	24%
Responsibility	17%	17%	10%	11%	–	10%
Self-policing	–	10%	37%	34%	44%	–
Accountability	13%	–	12%	14%	14%	38%
Miscellaneous	–	–	–	–	–	14%

(± 0.8) and 3.5 (± 0.9) respectively, $P = 0.04$]. However, there was no significant gender difference in the number of negative comments [mean (± sd) of 1.8 (± 1.5) and 1.9 (± 1.6), males and females respectively, $P = 0.85$].

Midpoint and End-of-course Evaluations

Students made more positive and negative comments (both peer- and self-provided) at the midpoint of the course than at the conclusion (Table 1). For both males and females, the mean (± sd) decrease in number of comments was as follows: positive self [0.4 (± 1.2), $P < 0.001$], positive peer [0.1 (± 1.6), $P = 0.22$] and negative peer [0.8 (± 1.1), $P < 0.001$].

Numerical ratings increased from the midpoint to the conclusion of the course, in both self-rating [mean of 3.5 and 3.6 respectively, for a mean (± sd) increase of 0.1 (± 1.0), $P = 0.05$] and peer ratings [mean of 4.2 (± 0.6) and 4.4 (± 0.5) respectively, for a mean (± sd) increase of 0.2 (± 0.6), $P < 0.001$].

Grading System

Despite the fact that the medical school switched from numerical grading to a pass/fail system, no significant difference was observed with respect to the number of positive comments (sum of peer and self comments; $P = 0.42$) or distribution of final grade percentage points ($P = 0.99$).

Discussion

Our results confirm those of Rudy et al[31] in that peer-provided positive comments correlated with higher peer evaluation scores, but not with self-evaluation scores. Additionally, we reproduce findings that students with higher grades underestimated their own performance, while those doing poorly tended to overestimate it.[32] These data suggest that, similar to previously drawn conclusions, peer assessment skills may not transfer to self-assessment skills.[23,31] Despite students' willingness to evaluate their peers in an objective fashion, their self-evaluation is much more critical. This may reflect students' reluctance to praise themselves, while being more comfortable commending their peers. It may also reflect a fear of an inflated self-rating, resulting in social isolation or a lower final grade (particularly in the face of a lower peer evaluation). It is also possible that lack of experience using this type of evaluation and unrealistic expectations of their own abilities as first-year students[31] may contribute to this observed difference.

The gender differences we demonstrate in peer evaluation and self-evaluation are consistent with those previously reported. Rees[33] reported that 73.3% of first-year male medical students overestimated their performance in professional development portfolios, while 72.7% of female students underestimated themselves. In a study of accuracy of third-year medical students' self-assessment, Lind et al[34] indicated that male students overestimated their performance on a surgery clerkship, while female students underestimated their abilities on the same rotation, despite the fact that female students were statistically outperforming male students. In our study, males may have received more positive peer-provided comments because they contributed more actively to group dissection. They may also have been more dominating in their groups. Nevertheless, these results indicate that gender equality in professional perception remains a goal to be actively pursued.

With respect to midpoint and end-of-course evaluations, the number of both positive and negative comments significantly decreased toward the end of the course. This may suggest that students had less interest in commenting on personal and peer behaviour toward the end of the

course. However, this may also indicate a decrease in unprofessional behaviour, as supported by the fact that the numerical rankings of professionalism increased toward the conclusion of the course.

The distribution of final grade percentages suggests that the grading system does not influence professionalism or students' motivation to achieve high marks in a course.[35]

The highlights of our findings can be summarised as follows:

- First-year medical students recognised professional values and commented on them during peer evaluation and self-evaluation.
- Peer assessment skills may not translate to self-assessment skills (the latter being more critical).
- Males often tended to overestimate their performance while females tended to underestimate theirs.
- Students performing better academically were more likely to provide positive comments about themselves and about their peers.
- Grading system does not appear to influence students' level of professionalism or motivation to achieve.

Conclusions

Professionalism is fundamental to medical practice and must be emphasized during medical education. However, despite the fact that medical students understand and are willing to provide peer assessment and self-assessment of professional behaviour, our results correlate with previously reported data to indicate that, at this point in their training, students lack the necessary insight to make accurate evaluations.[29] We submit that, rather than using peer evaluation and self-evaluation as a precise estimation of professionalism, such assessment may instead be applied as a training tool to help students learn to realistically appraise their own and colleagues' professional behaviour. When introduced early in the curriculum, such assessment may ultimately improve the accuracy of peer evaluation and self-evaluation conducted later in medical school.[33,34] Further studies are warranted to elucidate how students internalise the professional behaviours they will take with them into their future careers.

Acknowledgements

The authors wish to thank Dr Esther H Rodriguez and Dr Sarah M Jump for their constructive comments and help with editing the manuscript. In addition, the authors extend their thanks to the students of Mayo Clinic College of Medicine for their participation in the peer evaluation and self-evaluation in the Gross and Developmental Anatomy Course. Finally, we would also like to thank William S Harmsen from the Mayo Clinic Department of Health Science Research, Division of Biostatistics for all his statistical work and Ruth L Pedersen for her secretarial help. Approval of this study was granted by the Mayo Foundation Institutional Review Board (protocol # 232-00).

Appendix 1. Students' peer and self-evaluation questionnaires available on the Gross and Developmental Anatomy Course website. The window for open-ended questions did not constrain the length of typed-in text. The same questionnaire was used at the midpoint and conclusion of the course.

Student Peer Evaluation

1. **Please rate each person in your dissection team using the scale below:**
   (1) Would have learned more if not in group
   (2) Average contribution
   (3) Neither helped nor hurt my learning
   (4) Contributed more than average
   (5) Truly outstanding at helping me learn
2. **Identify an incident(s) that this team member was most helpful:**
3. **Identify an incident(s) that this team member could have handled differently:**

Student Self-evaluation

1. **Please rate your contribution to your assigned dissection team using the scale below:**
   (1) I made little or no contribution
   (2) Average contribution
   (3) Neither helped nor hurt others learning
   (4) Contributed more than average
   (5) Truly outstanding at helping others learning
2. **Identify an incident(s) in which you were effectively interacting with your team members:**
3. **Identify an incident(s) that you could have handled better or you could have been more effective in interacting with members of your team:**

REFERENCES

1. ABIM American Board of Internal Medicine Foundation; ACP-ASIM Foundation American College of Physicians-American Society of Internal Medicine; European Federation of Internal Medicine. Medical professionalism in the new millennium: a physician charter. Ann Intern Med 2002;136:243-6.
2. Patenaude J, Niyonsenga T, Fafard D. Changes in students' moral development during medical school: a cohort study. CMAJ 2003;168: 840-4.
3. Frader J, Arnold R, Coulehan J, Pinkus RL, Meisel A, Schaffner K. Evolution of clinical ethics teaching at the University of Pittsburgh. Acad Med 1989;64:747-50.
4. Sullivan WM. What is left of professionalism after managed care? Hastings Cent Rep 1999;29:7-13.
5. Hafferty FW, Franks R. The hidden curriculum, ethics teaching, and the structure of medical education. Acad Med 1994;69:861-71.
6. Barnitt R. Deeply troubling questions: the teaching of ethics in undergraduate courses. Br J Occup Ther 1993;56:401-6.
7. Crandall SJ, Volk RJ, Loemker V. Medical students' attitudes toward providing care for the underserved. Are we training socially responsible physicians? JAMA 1993;269:2519-23.
8. Shorr AF, Hayes RP, Finnerty JF. The effect of a class of medical ethics on first-year medical students. Acad Med 1994;69:998-1000.
9. Hebert PC, Meslin EM, Dunn EV. Measuring the ethical sensitivity of medical students: a study at the University of Toronto. J Med Ethics 1992;18:142-7.
10. Self DJ, Wolinsky FD, Baldwin DC Jr. The effect of teaching medical ethics on medical students' moral reasoning. Acad Med 1989;64:755-9.
11. Self DJ, Schrader DE, Baldwin DC Jr, Wolinsky FD. The moral development of medical students: a pilot study of the possible influence of medical education. Med Educ 1993;27:26-34.
12. Self DJ, Olivarez M, Baldwin C. Moral Reasoning in Medicine. Hillsdale (NJ): University of Minnesota Press, 1994.
13. Osborn E. Punishment: a story for medical educators. Acad Med 2000;75:241-4.
14. Self DJ, Baldwin DC Jr. Does medical education inhibit the development of moral reasoning in medical students? A cross-sectional study. Acad Med 1998;73:S91-S93.
15. O'Connell MT, Pascoe JM. Undergraduate medical education for the 21st century: leadership and teamwork. Fam Med 2004;36:S51-S56.
16. Swick HM, Szenas P, Danoff D, Whitcomb ME. Teaching professionalism in undergraduate medical education. JAMA 1999;282:830-2.
17. Prislin MD, Lie D, Shapiro J, Boker J, Radecki S. Using standardized patients to assess medical students' professionalism. Acad Med 2001;76:S90-S92.
18. Papadakis MA, Loeser H, Healy K. Early detection and evaluation of professionalism deficiencies in medical students: one school's approach. Acad Med 2001;76:1100-6.
19. Papadakis MA, Osborn EH, Cooke M, Healy K; University of California, San Francisco School of Medicine Clinical Clerkships Operation Committee. A strategy for the detection and evaluation of unprofessional behavior in medical students. Acad Med 1999;74:980-90.
20. Hemmer PA, Hawkins R, Jackson JL, Pangaro LN. Assessing how well three evaluation methods detect deficiencies in medical students' professionalism in two settings of an internal medicine clerkship. Acad Med 2000;75:167-73.
21. Epstein RM, Hundert EM. Defining and assessing professional competence. JAMA 2002;287:226-35.
22. Gordon J. Assessing students' personal and professional development using portfolios and interviews. Med Educ 2003;37:335-40.
23. Arnold L. Assessing professional behavior: yesterday, today, and tomorrow. Acad Med 2002;77:502-15.
24. Kao A, Lim M, Spevick J, Barzansky B. Teaching and evaluating students' professionalism in US medical schools, 2002-2003. JAMA 2003;290:1151-2.
25. Swick HM. Toward a normative definition of medical professionalism. Acad Med 2000;75:612-6.
26. Hensel WA, Dickey NW. Teaching professionalism: passing the torch. Acad Med 1998;73:865-70.
27. Cruess SR, Cruess RL. Professionalism must be taught. BMJ 1997;315:1674-7.
28. McDowell L, Mowl G. Innovative assessment – its impact on students. In: Gibbs G, editor. Improving Student Learning Through Assessment and Evaluation. Oxford: The Oxford Centre for Staff Development, 1996.
29. Rees C, Shepherd M. Students' and assessors' attitudes towards students' self-assessment of their personal and professional behaviours. Med Educ 2005;39:30-9.
30. Carmichael SW, Pawlina W. Loose lips sink ships. Acad Med 2004; 79:1002.
31. Rudy DW, Fejfar MC, Griffith CH 3rd, Wilson JF. Self- and peer assessment in a first-year communication and interviewing course. Eval Health Prof 2001;24:436-45.
32. Edwards RK, Kellner KR, Sistrom CL, Magyari EJ. Medical student self-assessment of performance on an obstetrics and gynecology clerkship. Am J Obstet Gynecol 2003;188:1078-82.
33. Rees C. Self-assessment scores and gender. Med Educ 2003;37:572-3.
34. Lind DS, Rekkas S, Bui V, Lam T, Beierle E, Copeland EM. Competency-based student self-assessment on a surgery rotation. J Surg Res 2002;105:31-4.
35. Jones KG, Pedersen RL, Carmichael SW, Pawlina W. Effects of pass/fail grading system on academic performance of first year medical students in the gross anatomy course. FASEB J 2003;17:A385.