

Differences in medical student self-evaluations of clinical and professional skills

This article was published in the following Dove Press journal:
Advances in Medical Education and Practice

Antoinette C Spoto-Cannons¹
Deanna M Isom²
Max Feldman³
Kira K Zwygart³
Rahul Mhaskar⁴
Marna Rayl Greenberg⁵

¹Department of Pediatrics, Morsani College of Medicine, University of Florida, Tampa, FL, USA; ²Department of Education, Morsani College of Medicine, University of Florida, Tampa, FL, USA; ³Department of Family Medicine, Morsani College of Medicine, University of Florida, Tampa, FL, USA; ⁴Department of Internal Medicine, Morsani College of Medicine, University of Florida, Tampa, FL, USA; ⁵Department of Emergency and Hospital Medicine, Lehigh Valley Campus, Morsani College of Medicine, University of South Florida, Bethlehem, PA, USA

Background: The skill of self-assessment is critical to medical students. We sought to determine whether there were differences between student self-assessments and their faculty assessments and if they were modified by gender. Additionally, we sought to determine the differences in these assessments between students in a traditional (core) versus an enhanced (SELECT) medical school curriculum.

Methods: In this retrospective study, mid-term and final assessment and feedback forms from the first-year Doctoring 1 course were analyzed from three academic years: 2014–2015 through 2016–2017. Data were abstracted from the forms and de-identified for analysis. Class year, student gender, and class type were also abstracted from this “on the shelf” data from program assessment. The level of agreement between faculty and student assessments was investigated using Wilcoxon signed ranks test. The gender differences (male versus female students) between student assessments and their assessment by their faculty were investigated by using the Kruskal Wallis test.

Results: Five hundred and thirty-five student self-assessments were analyzed. Fifty-six percent (301/535) were male while 44% (234/535) were female. Faculty assessments (P -value < 0.001) were higher than students and this was not modified by student gender. Compared to the domain of “participation” in the core program, there was no difference between the student/faculty ratings based on student gender (P -value: 0.48); there was a difference in the SELECT program cohort (P -value: 0.02). Specifically, the female students appear to rate themselves lower (female student: mean/standard deviation: 2.07/0.52) compared to their faculty (faculty: mean/standard deviation: 2.42/0.55).

Conclusion: Faculty consistently assessed the students at a higher rating than the students rated themselves. The level of difference between student self-assessments and their assessment by their faculty was not modified by student gender. With the minor exception of “participation,” there was no difference between students in the two different doctoring class curriculums.

Keywords: doctoring, self-evaluations, undergraduate medical, education

Introduction

Developing today’s medical students into successful leaders is an essential component of medical education.¹ Leadership is composed of several competencies that can be learned, including the self-confidence to lead.² This self-confidence comes from learning different aspects about oneself, such as skills and biases, self-awareness of weaknesses, and developing understanding about how these characteristics affect one’s relationships with others.² Self-assessment is a crucial part of medical education and plays an important role in establishing a physician’s competence.³ When evaluating self-assessment, there are variables that should be considered—for instance, the gender of the learner and the program nature.

Correspondence: Marna Rayl Greenberg
Department of Emergency and Hospital
Medicine, Lehigh Valley Hospital, Morsani
College of Medicine, University of South
Florida, LVH-M-5th Floor EM Residency
Suite, 2545 Schoenersville Road,
Bethlehem, PA 2545, USA
Email mrgdo@ptd.net

According to the American Association of Medical Colleges (AAMC), 50.7% of medical student matriculates were women in 2017.⁴ While there have been reports in a meta-analysis in which medical students' self-assessment and performance have been evaluated, 77% of those studied did not report gender.⁵ Additionally, in this same meta analysis of 35 articles, only 10 reported the self-assessment precision by gender and merely four had the sample described by gender composition, therein no comparison of effect sizes could be done.⁵ National Institutes of Health (NIH) implemented a policy recently [NOT-OD-15-102] to encourage scientists to report findings considering the possible part gender plays as a variable (SABV) in their work.⁶ Optimally, this policy will begin to impact education literature. Gender is a potential influence on self-assessment precision; research demonstrates that self-confidence is often an issue for female medical students.⁷ Additionally, female medical students compared to male medical students have reported greater stress about their abilities.⁸

"On Doctoring" (OD) courses in undergraduate medical education curriculum have a considerable internet presence even though relatively new.⁹ By searching for "on doctoring" courses (along with associated terms such as "art of medicine," "art of doctoring," and "physicianship") it can be seen that this type of curriculum has been adopted by numerous United States medical Schools.⁹ University of South Florida Morsani College of Medicine (USF MCOM) has an established Doctoring course that is unique in composition. Doctoring at USF MCOM is a small group-based sequence that teaches students interviewing, physical diagnosis, and differential diagnostic skills; bioethics, medical humanities, health systems and economics; community, preventive, and public health. It introduces care of special populations including the disabled. Our doctoring classes are split into two cohorts—Core and SELECT. The core program is a traditional doctoring curriculum. The SELECT doctoring curriculum is delivered as a part of an overarching SELECT program initiated in 2011 between the USF MCOM (Tampa, Florida) and the Lehigh Valley Health Network (LVHN, Allentown, PA). It promotes excellence in scholarly activities, experiences in leadership, and training that is collaborative. This program aims to prepare medical students with knowledge and skills to transform the nature of health care by providing training in leadership that is based on emotional intelligence.¹⁰ The same academic requirements are used for all students admitted to USFs, but students

admitted to this specific program are qualified based on the results of a secondary behavioral event interview designed to identify those with the greatest emotional intelligence and aptitude for leadership.¹⁰

To date, no gender-specific analysis has been performed on these program self-assessments to evaluate whether male or female students are similar in their likelihood to overrate or underrate themselves on these parameters or if there are differences between the core and SELECT students in their self-assessments. We set out to weave these concepts of leadership skill development and gender-specific self-assessment capability by comparing program outcomes in our doctoring course. In this study, we set forth the following hypotheses:

1. There is agreement between student self-assessments and faculty preceptor assessments.
2. The level of agreement between student self-assessments and faculty preceptor assessments is modified by student gender. Specifically, female medical students rate themselves lower than males in their self-assessments of their clinical and professional skills than their faculty preceptors do.
3. There will be a difference between Core and SELECT students in this potential gender gap.

With these theories in mind, we set out to assess whether there was a difference in agreement between medical students in self-assessments and their faculty preceptor rating of their clinical and professional skills as measured by their mid-term and final assessment and feedback forms.

Methods

Mid-term and final assessment and feedback forms ([Supplementary 1](#) and [2](#)) are obtained as part of the first-year longitudinal Doctoring 1 course. The mid-term feedback form is completed at the end of the first semester and the final assessment is completed the last day of the course during their second semester. Both forms are first completed by the student in order to promote self-reflection on their clinical skills (history taking skills; physical exam skills; communication skills; and patient-centered care) and professionalism (punctuality, timeliness; response to feedback; prepared for small group; participation in small group; and respect for patients) asking them to rate themselves as needs work, on target, or above average. The student then meets face-to-face with their preceptor, and after having the student self-reflect, the preceptor gives

verbal feedback, completes the form using the same rating scale, and submits it to the course coordinator and director.

After Institutional Review Board (IRB) approval, this retrospective, single university cohort study was conducted through the coordinated efforts of the USF MCOM (Tampa, FL) and Lehigh Valley (Allentown, PA) campus research study team members. Mid-term and final assessment and feedback forms from the first-year Doctoring 1 course were sorted by the course coordinator from the three academic years 2014–2015, 2015–2016, and 2016–2017 and evaluations were de-identified for analysis. The three years were a convenience sample of available evaluations and all available evaluations from students taking the doctoring (core or select) course were included. The two primary reviewers who extracted the data received training on how the data were to be interpreted. Students in both arms of this training program were assessed at mid-term and end-of-year and were provided feedback. Faculty had development sessions as a part of standard program work to prepare them to accurately perform their assessment tasks. No inter-rater reliability testing was performed between faculty. Class year, student gender, core or SELECT class, were also abstracted from this “on the shelf” data from program assessment.

For the sample size calculations; agreement between student and preceptor assessment was evaluated using the approach outlined by Bland and Altman,¹¹ where we compared 95% confidence levels of agreement, with a relative difference of >20% being considered significant. The sample size in this case (540) was incumbent on the width of the confidence intervals for the groups. Using a sample of 100 is recommended by Bland-Altman which yields a 95% CI of ± 0.34 of the standard deviation. This is equivalent to 1/6 the size of the definite agreement between students and preceptors. The power to detect a difference of greater than 20% is 80% if:

$$\frac{\delta + 2\sigma - \gamma}{\sqrt{\frac{3\sigma^2}{n}}} < -2.486,$$

where γ equals a 20% limit and δ and σ^2 are correspondingly the expected value and the variance of the difference (d). To achieve the recommended sample size,¹¹ we planned to use at least 100 pairs of student and preceptor evaluations.

The level of agreement between faculty and student assessments was investigated using Wilcoxon signed ranks test. The gender differences (male versus female students)

between student assessments and their assessment by their faculty were investigated by using the Kruskal Wallis test. These analyses were also conducted separately for students in the CORE versus the SELECT MD program to investigate whether there are any differences in the outcomes comparing these two student cohorts. We did not use imputation methods to account for missing data. The amounts of missing data were similar across student versus faculty and hence did not impact our overall analyses and findings. SPSS software version 24 was used to conduct all statistical analyses and $P < 0.05$ was considered significant.

Results

There are 535 students in our study cohort. Seventy percent (377/535) were from the core program and 30% (158/535) were from the SELECT program. Fifty-six percent (301/535) of our students were male while 44% (234/535) were females. Moreover, in the core program; 58% (220/377) of students were male while in the SELECT program 51% (81/158) of students were male.

Overall, the faculty consistently ranked the students at a higher rating than the students for all the assessment domains/questions (Table 1) during mid-term and the end-of-course evaluations. Overall, the level of difference between students and their faculty assessments was not modified by student gender during the end of the course of evaluations (Table 2). Similarly, mid-term evaluations showed no difference between student and faculty assessment based on student gender.

With only one exception, no differences were found by gender between the CORE and SELECT students at the end of the course evaluations. Compared to the domain of “participation” in the core MD program students where there was no difference between the student/faculty ratings based on student gender (P -value: 0.48), there was a difference in the SELECT MD program cohort (P -value: 0.02). Specifically, the female students appear to rate themselves lower (female student: mean/standard deviation: 2.07/0.52) compared to their faculty (faculty: mean/standard deviation: 2.42/0.55). Hence, the difference (mean difference/standard deviation: 0.33/0.47) between their rating and their faculty rating is higher compared with the difference (mean difference/standard deviation: 0.12/0.58) between their male colleagues’ rating (male student: mean/standard deviation: 2.20/0.54) and their faculty rating (faculty: mean/standard deviation: 2.34/0.50).

Table 1 Comparison between faculty and student assessments

Domain	Mid-term evaluations						End of course evaluations					
	N	Rankings	Mean	SD	Z	P-value (Wilcoxon signed ranks test)	N	Rankings	Mean	SD	Z	P-value (Wilcoxon signed ranks test)
History	469	Student	2.04	0.45	-8.72	<0.001	432	Student	2.11	0.42	-11.36	<0.001
	471	Faculty	2.28	0.48			434	Faculty	2.48	0.51		
Physical	463	Student	1.60	0.52	-12.64	<0.001	433	Student	2.03	0.45	-9.19	<0.001
	462	Faculty	1.99	0.33			434	Faculty	2.31	0.50		
Communication	468	Student	2.12	0.48	-7.93	<0.001	433	Student	2.22	0.50	-9.83	<0.001
	470	Faculty	2.36	0.52			433	Faculty	2.53	0.52		
Patient care	468	Student	2.20	0.47	-6.68	<0.001	433	Student	2.33	0.50	-9.69	<0.001
	466	Faculty	2.39	0.51			434	Faculty	2.62	0.49		
Punctuality	468	Student	2.26	0.47	-3.98	<0.001	432	Student	2.23	0.51	-7.48	<0.001
	470	Faculty	2.37	0.50			432	Faculty	2.45	0.54		
Feedback	467	Student	2.15	0.80	-9.56	<0.001	433	Student	2.19	0.43	-10.45	<0.001
	472	Faculty	2.44	0.50			431	Faculty	2.53	0.53		
Preparation	468	Student	2.05	0.38	-11.48	<0.001	433	Student	2.11	0.41	-11.48	<0.001
	472	Faculty	2.40	0.51			432	Faculty	2.47	0.52		
Participation	468	Student	2.13	0.54	-9.84	<0.001	433	Student	2.23	0.52	-9.95	<0.001
	472	Faculty	2.43	0.60			432	Faculty	2.54	0.52		
Respect	468	Student	2.34	0.47	-6.83	<0.001	432	Student	2.41	0.49	-9.05	<0.001
	470	Faculty	2.54	0.49			431	Faculty	2.69	0.46		

Table 2 Impact of student gender on end of the course self-assessments

Domain	Gender	N	Mean Rank	Chi-Square	P-value (Kruskal Wallis test)
History	Male	244	206.43	1.36	0.24
	Female	178	218.45		
Physical	Male	244	204.60	2.92	0.09
	Female	179	222.09		
Communication	Male	243	205.63	1.82	0.18
	Female	179	219.46		
Patient care	Male	244	209.35	0.38	0.54
	Female	179	215.61		
Punctuality	Male	243	208.53	0.21	0.65
	Female	177	213.21		
Feedback	Male	243	209.57	0.04	0.83
	Female	177	211.77		
Preparation	Male	243	209.29	0.16	0.69
	Female	178	213.34		
Participation	Male	243	204.05	2.57	0.11
	Female	178	220.49		
Respect	Male	242	205.41	1.35	0.24
	Female	178	217.42		

Discussion

In our study, faculty consistently assessed the students at a higher rating than the students ranked themselves and the level of difference between student self-assessments and their faculty assessments was not overall modified by student gender. This contrasts with a 2010 meta-analysis⁶ which found that students were moderately able to self-assess themselves and that female students underestimated their performance more than males. We had hoped to illustrate that the SELECT leadership curriculum narrowed any gender gap in self-assessments; however, with only one exception (in the domain of “participation”) we found no gender difference between the two doctoring curriculums (CORE and SELECT) in regard to ratings.

Past research⁷ reported that despite performing equally, female students consistently reported decreased self-confidence which should impact their self-assessments. Prior objectives for medical educators based on this earlier research were to focus on issues regarding female student’s confidence and perceptions.⁷ More recently, Madrazo (et al) found that females underestimated their performance in clinical exams.¹² More specific gender differences have been illustrated with regard to the interaction of anxiety (higher anxiety improved accuracy of self-assessment for females but not in males).¹³ We were pleased to see that there were no significant differences between genders.

It is noted that the primary hypotheses in this study were not substantiated by our study results. Findings in this study may be limited due to the evaluation of small sample size and a single site for a single course during their first year of medical school. Additionally, it is not known why faculty had higher assessments than the students’ self-assessments. Confounding variables such as prior training experience, faculty evaluator skills (and inter-rater reliability), age of students and faculty were not controlled for since this was ‘on the shelf’ data. Future studies may look at the self-assessment of medical students through all four years. This will not only provide a look into whether gender difference affects self-confidence but may give a better glance into the effect of the SELECT leadership curriculum.

Conclusion

Faculty consistently assessed the students at a higher rating than the students rated themselves. The level of

difference between student assessments of their own performance and their assessments by their faculty was not modified by student gender. With the minor exception of “participation,” there was no difference between students in the two different doctoring class curriculums.

Acknowledgment

Authors would like to acknowledge the medical editing assistance of Lexis Laubach, BS.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Simpson J. Making and preparing leaders. *Med Educ.* 2000;34:2110–2115.
2. Kouzes JM, Posner BZ. *The Leadership Challenge*. San Francisco: Jossey Bass, Inc; 1987.
3. Rees C, Shepherd M. Students’ and assessors’ attitudes toward students’ self-assessment of their personal and professional behaviours. *Med Educ.* 2005;39:30–39. doi:10.1111/j.1365-2929.2004.02030.x
4. Association of American Medical Colleges. More women than men enrolled in U.S. medical schools in 2017. Available from: <https://news.aamc.org/press-releases/article/applicant-enrollment-2017>. Accessed August 16, 2019.
5. Hartigan D. Medical students’ self-assessment of performance: results from here meta-analyses. *Patient Educ Couns.* 2011;84(1):3–9. doi:10.1016/j.pec.2010.06.037
6. Consideration of sex as a biological variable in NIH-funded Research. National Institutes of Health Office of Extramural Research. Available from: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-102.html>. Accessed February 09, 2017.
7. Blanch D, Hall J, Roter D, Frankel R. Medical student gender and issues of confidence. *Patient Educ Couns.* 2008;72:374–381. doi:10.1016/j.pec.2008.05.021
8. Moffat KJ, McConnachi A, Ross S, Morrison JM. First Year medical student stress and coping in a problem-based learning medical curriculum. *Med Educ.* 2004;38:482–491. doi:10.1046/j.1365-2929.2004.01814.x
9. Hafferty F, Gaufrberg E, O’Donnell J. The role of the hidden curriculum in “On Doctoring Courses”. *AMA J Ethics.* 2015;17(2):129–137. doi:10.1001/virtualmentor.2015.17.2.medu1-1502
10. Roscoe LA, English A, Monroe AD. Scholarly excellence, leadership experiences and collaborative training: qualitative results from a new curricular initiative. *J Contemp Med Educ.* 2014;2:163–167. doi:10.5455/jcme.20140928035359
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307–310. doi:10.1016/S0140-6736(86)90837-8
12. Madrazo L, Lee C, McConnell M, Khamisa K. Self-assessment differences between genders in a low-stakes objective structured clinical examination (OSCE). *BMC Res Notes.* 2018;11(393). doi:10.1186/s13104-018-3494-3
13. Colbert-Getz J, Fleishman C, Jung J, Shikofski N. How do gender and anxiety affect students’ self-assessment and actual performance on a high-stakes clinical skills examination? *Acad Med.* 2013;88:44–48. doi:10.1097/ACM.0b013e318276bcc4

Advances in Medical Education and Practice**Dovepress****Publish your work in this journal**

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education

including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>