

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/312503867>

Clinical Performance Evaluations of Third-Year Medical Students and Association With Student and Evaluator Gender

Article in *Academic Medicine* · January 2017

DOI: 10.1097/ACM.0000000000001565

CITATIONS

18

READS

234

5 authors, including:



Brian Alverson

Brown University

59 PUBLICATIONS 2,447 CITATIONS

[SEE PROFILE](#)



Sangshin Park

University of Seoul

90 PUBLICATIONS 929 CITATIONS

[SEE PROFILE](#)



Randal M Rockney

Brown University

36 PUBLICATIONS 290 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Sj vaccine [View project](#)



moms project [View project](#)

Clinical Performance Evaluations of Third-Year Medical Students and Association With Student and Evaluator Gender

Alison Riese, MD, MPH, Leah Rappaport, MD, Brian Alverson, MD, Sangshin Park, DVM, MPH, PhD, and Randal M. Rockney, MD

Abstract

Purpose

Clinical performance evaluations are major components of medical school clerkship grades. But are they sufficiently objective? This study aimed to determine whether student and evaluator gender is associated with assessment of overall clinical performance.

Method

This was a retrospective analysis of 4,272 core clerkship clinical performance evaluations by 829 evaluators of 155 third-year students, within the Alpert Medical School grading database for the 2013–2014 academic year. Overall clinical performance, assessed on a

three-point scale (meets expectations, above expectations, exceptional), was extracted from each evaluation, as well as evaluator gender, age, training level, department, student gender and age, and length of observation time. Hierarchical ordinal regression modeling was conducted to account for clustering of evaluations.

Results

Female students were more likely to receive a better grade than males (adjusted odds ratio [AOR] 1.30, 95% confidence interval [CI] 1.13–1.50), and female evaluators awarded lower grades than males (AOR 0.72, 95% CI 0.55–0.93), adjusting for department,

observation time, and student and evaluator age. The interaction between student and evaluator gender was significant ($P = .03$), with female evaluators assigning higher grades to female students, while male evaluators' grading did not differ by student gender. Students who spent a short time with evaluators were also more likely to get a lower grade.

Conclusions

A one-year examination of all third-year clerkship clinical performance evaluations at a single institution revealed that male and female evaluators rated male and female students differently, even when accounting for other measured variables.

Selection of graduating medical students into residency programs is driven by multiple factors. However, according to program directors, the most important selection criteria are students' grades on required core clerkships.¹ Clinical performance evaluations (CPEs) are used in most core clinical clerkships as assessment and grading tools for medical students. Clinicians who work with medical students are asked to complete formal evaluations of each student's basic clinical skills, such as history taking and case presentation, as well as fund of knowledge and professionalism. In most clerkships, these evaluations, along with standardized written examinations and objective structured clinical examinations (OSCEs), provide the data from which students' final clerkship grades are

determined. Studies show that these CPEs are weighted more heavily than the other evaluation methods, accounting for 50% to 70% of the final grade across all clerkships.^{2,3} Despite the importance of core clerkship clinical evaluations, there is a paucity of literature examining the degree of objectivity of this measure.⁴

The numerous evaluations that occur over the course of attaining entrance to medical school and during the preclinical years are largely standardized and unlikely to exhibit grader-dependent bias. In contrast, medical students are evaluated in a more subjective manner when being assessed on their clinical performance. For that reason, the association of grading with gender and the gender pairing of trainer and trainee is important, yet these factors are not well understood in the medical setting in areas where subjectivity of grading is high. Literature from the education field has shown that student gender often plays a role in how students are treated and graded.^{5,6} In primary schools, girls are awarded better grades than boys, despite similar test scores, which some

researchers attribute to “noncognitive skills”—specifically, “a more developed attitude towards learning.”⁶ Additionally, teachers' gender can affect their expectations and perceptions of educational competence and performance.^{7,8} Furthermore, studies^{9–11} suggest that gender pairing can enhance, through a “role-model effect,” student engagement and behavior, or, conversely, gender noncongruence may induce “stereotype threat,” in which anxiety that one will confirm a negative stereotype can lead to a decrement in performance.

A few small studies^{12–14} have suggested an interaction between student and evaluator gender in the grading of medical students' simulated clinical performance on OSCEs by standardized patients (SPs). One small study of OSCE grading¹³ found that male and female medical students fared similarly overall; however, when graded solely by female SPs, women scored significantly higher, yet male and female students were rated the same by male SPs. These findings were replicated in a more recent study of OSCE grading,¹⁴ which specifically

Please see the end of this article for information about the authors.

Correspondence should be addressed to Alison Riese, Department of Pediatrics, Alpert Medical School, Hasbro Children's Hospital, 593 Eddy St., Providence, RI 02903; telephone: (401) 444-8531; e-mail: ariese@lifespan.org.

Acad Med. 2017;92:835–840.

First published online January 17, 2017
doi: 10.1097/ACM.0000000000001565

examined the gender interaction during a “gender-sensitive” patient situation, the examination of the chest.

Similar disparities in grading regarding student and evaluator gender have been found in a few small studies of nonsimulated clinical settings.^{15,16} A small study of students completing a required one-month ambulatory care medicine clerkship at the Medical College of Wisconsin¹⁶ showed that the highest mean grade was given by male preceptors to female students, and the lowest mean grade was given by female preceptors to male students. In a study of evaluations of internal medicine residents, male residents received higher grades from male attendings than from female attendings.¹⁷ Conversely, a study of medical student grading in obstetrics–gynecology¹⁸ found that female students performed better on written exams and OSCEs; however, they were graded similarly to male students by their faculty evaluators.

The influence of gender on grading in the clinical setting is important to understand, considering the highly subjective nature of clinical evaluations compared with multiple-choice tests, where gender has no bearing on grade assignment, as well as the more structured setting of OSCEs, where graders are generally well trained and have more uniform interactions with the students being assessed. CPEs are completed by evaluators of all training levels, who interact with students in various types of settings and over varying durations, yet their assessments are weighted heavily in clerkship grading.

As a first step in any effort to increase objectivity in clinical grade assignment, it is necessary to fully understand what issues influence evaluators’ grading of student clinical performance. There has been no study examining third-year core clerkships as a whole to see how the gender of the evaluator and the gender of the student may be associated with differences in the clinical evaluation of the student. We carried out this study to determine whether student and evaluator gender is associated with the grades assigned. Secondly, we sought to explore other student and evaluator factors that may be associated with variance in grading.

Method

This was a retrospective study conducted at the Alpert Medical School (AMS). All 4,462 CPEs recorded in the medical school’s grading database (OASIS) from third-year core clerkships during the 2013–2014 academic year were initially included. At AMS, the core clerkships and their duration during the study period consisted of internal medicine (12 weeks) and surgery, obstetrics–gynecology, family medicine, pediatrics, and psychiatry (each 6 weeks). The medical school’s administrative offices compiled deidentified demographic information about the student and evaluator for each CPE and assigned an ID number for each student and each evaluator who was involved in the CPEs being studied. The evaluator IDs were used to account for nesting of evaluations among evaluators—that is, cases where evaluators assessed more than one student. As the indicator of the student’s global clinical performance, the “overall clinical performance” grade on the CPE, which from now on we refer to as the student *grade*, was extracted from each CPE. The possible grades that could be selected by each evaluator completing a CPE were “exceptional,” “above expectations,” “meets expectations,” and “below expectations.” An evaluation was excluded if it was noted to be a duplicate entry or if data were incomplete for the primary outcome or predictor variables. Additionally, CPEs with a grade of “below expectations” were excluded because of the rare occurrence (< 1%) of this grade.

Because we were provided deidentified data, we were not able to match those data with any objective nonclinical evaluations. However, we did compare the United States Medical Licensing Examination (USMLE) Step 1 scores for men versus women in the class of 2015. The medical school administrative offices provided the means and standard deviations (SDs) of the USMLE Step 1 scores for the male and female students in that class, since these students’ CPE data were in our study. The means and SDs for these two groups were compared using Student *t* test. This study was declared exempt by the Lifespan institutional review board. (Lifespan Corporation, Rhode Island’s largest health system, is affiliated with the AMS of Brown University.)

Table 1
Demographic Information Regarding the Third-Year Medical Students and Their Evaluators at Alpert Medical School, 2013 to 2014^a

Characteristic	No. (%)
Students (n = 155)	
Gender	
Male	76 (49.0)
Female	79 (51.0)
Age quartile	
25–26 years	49 (33.3)
27 years	36 (24.5)
28 years	27 (18.4)
> 28 years	35 (23.8)
Evaluators (n = 829)	
Gender	
Male	399 (48.1)
Female	430 (51.9)
Age quartile	
25–30 years	290 (37.8)
31–40 years	210 (27.4)
41–50 years	137 (17.9)
> 50 years	130 (17.0)
Training level	
Resident (PGY 1)	168 (22.3)
Resident (PGY 2)	89 (11.8)
Resident (PGY 3–5)	155 (20.6)
Attending	342 (45.4)
Department	
Family medicine	126 (15.2)
Psychiatry	50 (6.0)
Internal medicine	336 (40.5)
Obstetrics–gynecology	47 (5.7)
Pediatrics	155 (18.7)
Surgery	115 (13.9)

Abbreviation: PGY indicates postgraduate year.
^aThe authors carried out a retrospective analysis of 4,272 third-year clerkship clinical performance evaluations, involving the evaluators and the students whose data are in this table, to determine whether student and evaluator gender are associated with assessment of overall clinical performance.

For each CPE, the dataset contained demographic information about the clerkship context, the student, and the evaluator. Clerkship characteristics for each CPE consisted of the clerkship department and the length of observation time for the student/evaluator (either < 2 half-days or ≥ 2 half-days). Student demographic information included student gender and age (grouped as 25–27 years old and ≥ 28 years old). Evaluator variables were evaluator gender,

age (in quartiles), and training level (residency year or attending).

All statistical analyses were performed using SAS 9.4 (SAS Institute, Cary, North Carolina). A *P* value < .05 was considered to be statistically significant. This study examined the associations of final grade with gender and covariates using chi-square tests. Hierarchical ordinal regression modeling was conducted to examine the effects of student and evaluator characteristics on a student's grade ("exceptional," "above expectations," or "meets expectations"), adjusting for nonindependence, or "clustering," of evaluators who rated more than one student. Gender and covariates with a *P* value < .05 in the univariable model were incorporated into a multivariable regression model, which was built by the stepwise selection procedure. Variables that significantly reduced residual variance were retained in the final model. To avoid colinearity, phi coefficients were estimated for two independent variables. If high colinearity among variables was observed ($r > 0.6$), we selected the most relevant variable to the student's grade for multivariable modeling. Because of the small number of evaluations in family medicine and psychiatry, data from these specialties were combined for the multivariable modeling. After the main effects model was built, interaction terms were explored for significance.

Results

Of the 4,462 CPEs initially included in this study, 190 (0.043%) were excluded. Thirty-eight were excluded because they were duplicates, and 136 were excluded because of missing values in predictors of interest (student or evaluator gender; no. = 18) or in the outcome of interest (grade; no. = 118). In addition, 16 CPEs were excluded because of a "below expectations" grade. Thus, the final study dataset comprised 4,272 CPEs, which were completed by 829 evaluators regarding the performance of 155 students. The mean (SD) USMLE Step 1 score for the AMS class of 2015 was 221 (18.70) for women and 231 (18.98) for men (*P* = .0083). The median age of students was 27 years (interquartile range [IQR] 26–28 years); the median age of evaluators was 33 years (IQR 29–45 years). (See Table 1 for student

and evaluator demographics.) While the number of students rotating through each clerkship was consistent, the number of CPEs for each student varied by clerkship. The internal medicine clerkship evaluators completed 1,267 CPEs (30% of all CPEs), and the pediatrics clerkship evaluators completed 1,154 (27%), which means that these two clerkships contributed a large percentage of CPEs compared with the percentages contributed by the other four clerkships.

There was variability in the number of CPEs per student (median 27, IQR 6–39) and CPEs per evaluator (median 3, IQR 1–7). Each clerkship, student, and evaluator characteristic examined was associated with a statistically significant difference in the distribution of grades received. (See Table 2.)

In univariable models, all predictors were associated with the grade. Because of high correlation between faculty age

Table 2

Associations of Third-Year Clinical Performance Grades With Clerkship, Student, and Evaluator Characteristics, Alpert Medical School, 2013 to 2014^a

Characteristic	No. (%) of all clinical evaluations (n = 4,272)	No. (%) of grades by category			<i>P</i> value
		Meets expectations (n = 721)	Above expectations (n = 1,826)	Exceptional (n = 1,725)	
Clerkships					
Department					
<i>Obstetrics-gynecology</i>	602 (14.1)	136 (22.6)	271 (45.0)	195 (32.4)	< .001
<i>Pediatrics</i>	1,154 (27.0)	236 (20.5)	483 (41.9)	435 (37.7)	
<i>Psychiatry</i>	300 (7.0)	45 (15.0)	136 (45.3)	119 (39.7)	
<i>Family medicine</i>	369 (8.6)	52 (14.1)	158 (42.8)	159 (43.1)	
<i>Surgery</i>	580 (13.6)	68 (11.7)	268 (46.2)	244 (42.1)	
<i>Internal medicine</i>	1,267 (29.7)	184 (14.5)	510 (40.3)	573 (45.2)	
Observation time					
≤ 2 half-days	505 (11.8)	143 (28.4)	261 (51.8)	100 (19.8)	< .001
> 2 half-days	3,785 (88.2)	578 (15.3)	1,565 (41.5)	1,625 (43.1)	
Students					
Gender					
Male	2,036 (47.7)	381 (18.7)	857 (42.1)	798 (39.2)	.009
Female	2,236 (52.3)	340 (15.2)	969 (43.3)	927 (41.5)	
Age					
25–27 years	2,542 (63.7)	482 (19.0)	1,091 (42.9)	969 (38.1)	< .001
≥ 28 years	1,448 (36.3)	211 (14.6)	617 (42.6)	620 (42.8)	
Evaluators					
Gender					
Male	2,081 (48.7)	263 (12.6)	893 (42.9)	925 (44.5)	< .001
Female	2,191 (51.3)	458 (20.9)	933 (42.6)	800 (36.5)	
Age quartile					
25–30 years	1,671 (41.2)	300 (18.0)	659 (39.4)	712 (42.6)	< .001
31–40 years	1,111 (27.4)	193 (17.4)	450 (40.5)	468 (42.1)	
41–50 years	701 (17.3)	88 (12.6)	325 (46.4)	288 (41.1)	
≥ 51 years	573 (14.1)	93 (16.2)	276 (48.2)	204 (35.6)	
Training level					
Resident (PGY 1)	763 (18.8)	115 (15.1)	281 (36.8)	367 (48.1)	< .001
Resident (PGY 2)	647 (15.9)	94 (14.5)	278 (43.0)	275 (42.5)	
Resident (PGY 3–5)	883 (21.7)	179 (20.3)	350 (39.6)	354 (40.1)	
Attending	1,776 (43.7)	287 (16.2)	826 (46.5)	663 (37.3)	

Abbreviation: PGY indicates postgraduate year.

^aThe authors carried out a retrospective analysis of 4,272 third-year clerkship clinical performance evaluations by 829 evaluators of 155 students to determine whether student and evaluator gender are associated with assessment of overall clinical performance.

Table 3

Odds Ratios of Receiving a Higher Grade—by Clerkship, Student, and Evaluator Characteristics—on 4,272 Third-Year Clerkship Clinical Performance Evaluations, Alpert Medical School, 2013 to 2014^{a,b}

Characteristic	Univariable model		Multivariable model	
	OR (95% CI)	P value	AOR (95% CI)	P value
Clerkships				
Department				
Obstetrics–gynecology	1.00 ^c	< .001	1.00 ^c	.002
Pediatrics	1.11 (0.67–1.83)		0.83 (0.49–1.40)	
Family medicine and psychiatry ^d	1.52 (0.91–2.53)		1.24 (0.71–2.17)	
Surgery	1.88 (1.10–3.21)		1.31 (0.74–2.33)	
Internal medicine	2.09 (1.30–3.36)		1.61 (0.97–2.69)	
Observation time				
≤ 2 half-days	1.00 ^c	< .001	1.00 ^c	< .001
> 2 half-days	2.72 (2.20–3.38)		2.74 (2.17–3.46)	
Students				
Gender				
Male	1.00 ^c	.016	1.00 ^c	.026
Female	1.17 (1.03–1.33)		1.30 (1.13–1.50)	
Age group				
25–27 years	1.00 ^c	< .001	1.00 ^c	< .001
≥ 28 years	1.39 (1.21–1.60)		1.48 (1.27–1.71)	
Evaluators				
Gender				
Male	1.00 ^c	< .001	1.00 ^c	.012
Female	0.64 (0.51–0.81)		0.72 (0.55–0.93)	
Age group				
25–30 years	1.00 ^c	.034	1.00 ^c	.028
31–40 years	0.78 (0.57–1.06)		0.76 (0.56–1.04)	
41–50 years	0.77 (0.54–1.09)		0.72 (0.50–1.03)	
≥ 51 years	0.59 (0.41–0.85)		0.57 (0.39–0.84)	
Training level ^e				
Attending	1.00 ^c	< .001	—	—
Resident (PGY 1)	1.90 (1.40–2.59)		—	—
Resident (PGY 2)	1.68 (1.21–2.35)		—	—
Resident (PGY 3–5)	1.29 (0.95–1.76)		—	—

Abbreviations: PGY indicates postgraduate year; OR, odds ratio; AOR, adjusted odds ratio.

^aThe authors carried out a retrospective analysis of 4,272 third-year clerkship clinical performance evaluations by 829 evaluators of 155 students to determine whether student and evaluator gender are associated with assessment of overall clinical performance.

^bA total of 32.9% of the variability in the grades was accounted for by “within-evaluator” nesting of grades in the multivariable model (intraclass correlation coefficient = 0.33; $P < .001$).

^cReference.

^dBecause of the small number of evaluations, family medicine and psychiatry were combined for the multivariable modeling.

^eTraining level was not retained in the multivariable model because of high phi coefficient (0.84) between faculty age quartile and training levels.

and training level (phi coefficient 0.84), only evaluator age was considered for the multivariable model. A total of 32.9% of the variability in the grades was accounted for by within-evaluator nesting of grades in the multivariable model (intraclass correlation coefficient = 0.329; $P < .001$). All significant differences in the univariable models were

retained in the multivariable model. In the multivariable model, female student gender was associated with higher grades (adjusted odds ratio [AOR], 1.30; 95% CI, 1.13–1.50). Female faculty gender was associated with lower grades (AOR, 0.72; 95% CI, 0.55–0.93). Longer observation time, older student age, and younger evaluator age were

all associated with higher grades. Evaluators in internal medicine had the highest odds of giving a better grade, while those in obstetrics–gynecology had the lowest odds. (See Table 3.)

The interaction between student and faculty gender, adjusted for all other main effects, was also significant ($P = .03$; see Figure 1). Male evaluators did not significantly differ in their grading of male and female students ($P = .29$); however, female evaluators gave lower grades to male students compared with female students ($P < .001$). Additionally, a significant interaction between faculty age and faculty gender was found ($P = .047$), with older male evaluators giving significantly lower grades than younger men ($P = .001$), while there was no significant difference in grading for the female age groups ($P = .71$). (See Figure 2). There was no interaction between student gender and student age ($P = .63$).

Discussion

In one year at a large U.S. medical school, there were over 4,000 CPEs of students in core clerkships, and data revealed that in clerkship grading, overall, male students received lower grades than female students on their CPEs. This finding is in accordance with literature examining gender differences in clinical performance. In general, male and female medical students perform similarly on the MCAT exam and have similar preclinical GPAs and USLME tests scores,^{15,19,20} albeit with factors including content area and student and school characteristics playing a role in performance. The class of students represented in our dataset actually differed in their performance on USMLE Step 1, with men performing significantly better. In contrast, other studies^{15,21} suggest that female medical students do tend to perform better on OSCEs, including those that are part of the USMLE Step 2 Clinical Skills (CS) exam, and receive better evaluations on their actual clinical performance. There was no interaction between evaluator gender and student gender found in the study of Step 2 CS scoring.²¹ However, our findings show that the discrepancy in clinical performance grades between male and female medical students was driven primarily by female evaluators.

The discrepancy between male and female evaluators' assessment of medical

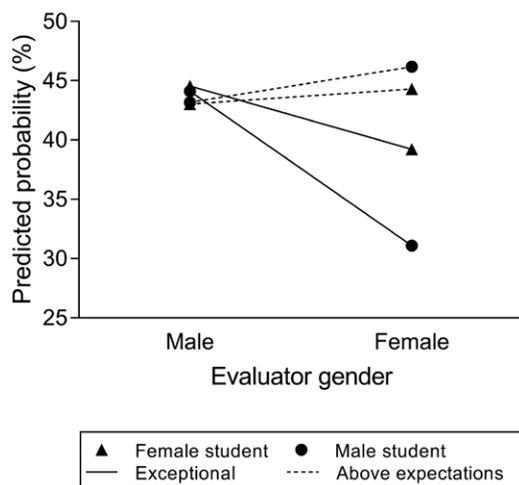


Figure 1 Two-way interaction effect of student and evaluator gender on predicted probabilities of evaluators' grading of male and female students on 4,272 clinical performance evaluations of medical students completing third-year core clerkships at Alpert Medical School, 2013 to 2014. Male evaluators did not significantly differ in their grading of male and female students ($P = .29$); however, female evaluators gave lower grades to male students than to female students ($P < .001$).

students' clinical performance is most perplexing. Medical students' clinical performance is influenced by attributes outside of medical knowledge and clinical acumen. Indeed, two studies^{22,23} reported that medical students who showed empathy received better clinical evaluations, and women scored higher on empathy scales than men did. Additionally, some studies^{22,23} found that female students' interpersonal skills surpassed those of men. In primary care, a study²⁴ showed that female physicians' communication skills surpassed those of their male counterparts, which, if

future studies confirm this result, is an important finding because doctor–patient communication has been linked to improved health outcomes.²⁵ If the body of literature showing that women outperform men in the clinical setting is applied, our findings suggest that female evaluators accurately detected superior performance in their female students, while male evaluators either were unable to detect these differences or were biased in their grading methods.

However, it is likely that this finding highlights an even more complicated

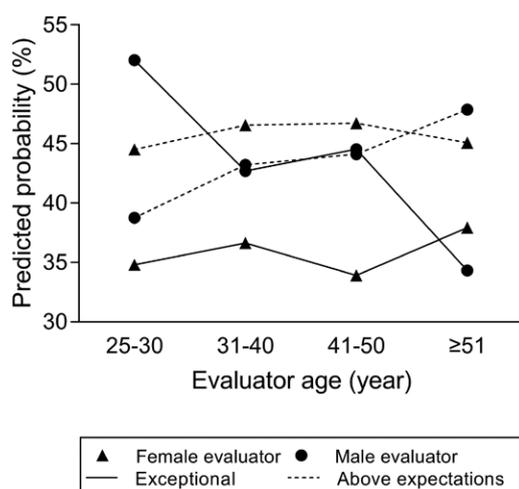


Figure 2 Two-way interaction effect of evaluator gender and age quartile on predicted probabilities of their grading of male and female students on 4,272 clinical performance evaluations of medical students completing third-year core clerkships at Alpert Medical School, 2013 to 2014. A significant interaction between faculty age and faculty gender was found ($P = .047$), with older male evaluators giving significantly lower grades than younger men ($P = .001$), while there was no significant difference in grading for the female age groups ($P = .71$).

interplay between gender and academic performance and assessment. As in the primary education world, female students' "learning attitude" may also play a role, as well as the possible role modeling of same-gender evaluators and the stereotype threat of opposite-gender graders, which may influence students to perform differently depending on the gender of their evaluators. Another potential complicating matter is that patients may interact differently with medical students depending on the student's gender, which could also affect the assessment of their performance. This has been demonstrated in a study²⁶ examining physician–patient interaction, where patients were found to speak differently and make more psychosocial disclosures to female physicians. Whatever the cause, it is concerning that our study findings suggest that male and female students experience different gradings of their clinical performances, and that the gender of the evaluator is an independent driver of this difference.

Our data also found a significant interaction between evaluator age and gender, with younger male evaluators awarding higher grades than older male evaluators and than female evaluators in all age groups. While younger evaluators have been found to be more lenient graders in other studies,^{27,28} to our knowledge the age–gender interaction has not been examined elsewhere, and this finding warrants additional investigation. Again, it is concerning that intrinsic evaluator characteristics have led to differential grading of students. Either improved training of graders is needed, or the characteristics of the evaluators must be taken into account when considering their ability to give fair clerkship grades.

Our data also demonstrate substantial differences in the way clerkship students are graded by department at our school, a finding that we suspect applies to many schools. This variability should be examined to provide a consistent approach to CPEs. Differences in the structure and duration of the different core clerkships, as well as the time students spend with evaluators, must be taken into consideration when looking at CPEs. In some cases, the structure of the clerkship and number of evaluators providing CPEs may result in fewer

grading events per student, which may exaggerate the influence of gender and age on a student's final clerkship grade.

Our study has some limitations. We evaluated only one year of grading events at one medical school in the United States. A multicenter study would be needed to see if these data are generalizable to other institutions. The grading system used is an ordinal one, and these data may not be reflective of data produced by other grading systems at other medical schools. We were not able to adjust for or compare clinical performance grades with standardized test scores, since the individual-level data were not available in our dataset. Further, we recognize that gender representation, and thus gender interactions at a medical school in 2013–2014, might be very different from what was obtained in previous years, when gender relationships and generational differences would perhaps skew data in other ways.

Further study is needed to learn whether the trends of gender-pairing influence on grading at our medical school are found at other medical schools. Additionally, the cause of the grading differences by evaluator and student gender is still unknown. Next steps may include a qualitative approach to discover reasons for the discrepancy in how medical students' performance is perceived and assessed by evaluators of different genders.

Acknowledgments: The authors would like to acknowledge the assistance of Alpert Medical Schools' Medical School Administrative Office for the compilation of the dataset used for this study. They would also like to thank Jennifer F. Friedman, MD, MPH, PhD, for her mentorship and guidance, and Kelvin Moore for his efforts assisting with the literature review.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: This study was reviewed by the Lifespan institutional review board and deemed exempt. (Lifespan Corporation, Rhode Island's largest health system, is affiliated with the Alpert Medical School of Brown University.)

Previous presentations: Pediatric Hospital Medicine Conference, Chicago, Illinois, July 29, 2016; and Lifespan Annual Research Celebration, Providence, Rhode Island, October 20, 2016.

A. Riese is assistant professor, Department of Pediatrics and Medical Science, Section of Medical Education, Alpert Medical School of Brown

University, Providence, Rhode Island.

L. Rappaport is a first-year pediatrics resident, University of Michigan Medical School, Ann Arbor, Michigan.

B. Alverson is associate professor, Department of Pediatrics and Medical Science, Section of Medical Education, Alpert Medical School of Brown University, Providence, Rhode Island.

S. Park is postdoctoral research associate, Alpert Medical School of Brown University and Center for International Health Research at Rhode Island Hospital, Providence, Rhode Island.

R.M. Rockney is professor, Department of Pediatrics, Family Medicine, and Medical Science, Section of Medical Education, Alpert Medical School of Brown University, Providence, Rhode Island.

References

- Green M, Jones P, Thomas JX Jr. Selection criteria for residency: Results of a national program directors survey. *Acad Med.* 2009;84:362–367.
- Kassebaum DG, Eaglen RH. Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Acad Med.* 1999;74:842–849.
- Hemmer PA, Papp KK, Mechaber AJ, Durning SJ. Evaluation, grading, and use of the RIME vocabulary on internal medicine clerkships: Results of a national survey and comparison to other clinical clerkships. *Teach Learn Med.* 2008;20:118–126.
- Holmboe ES. Faculty and the observation of trainees' clinical skills: Problems and opportunities. *Acad Med.* 2004;79:16–22.
- Lavy V, Sand E. On the Origins of Gender Human Capital Gaps: Short- and Long-Term Consequences of Teachers' Stereotypical Biases. Cambridge, MA: National Bureau of Economic Research; 2015.
- Cornwell C, Mustard DB, Van Parys J. Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *J Hum Resour.* 2013;48:236–264.
- Mullola S, Ravaja N, Lipsanen J, et al. Gender differences in teachers' perceptions of students' temperament, educational competence, and teachability. *Br J Educ Psychol.* 2012;82(pt 2):185–206.
- Heyder A, Kessels U. Do teachers equate male and masculine with lower academic engagement? How students' gender enactment triggers gender stereotypes at school. *Soc Psychol Educ.* 2015;18:467–485.
- Dee TS. Teachers and the gender gaps in student achievement. *J Hum Resour.* 2007;42:528–554.
- Keller J. Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *Br J Educ Psychol.* 2007;77(pt 2):323–338.
- Huguet P, Regner I. Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *J Educ Psychol.* 2007;99(3):545.
- Ramsbottom-Lucier M, Johnson MM, Elam CL. Age and gender differences in students' preadmission qualifications and medical school performances. *Acad Med.* 1995;70:236–239.
- Dawson-Saunders B, Rutala PJ, Witzke DB, Leko EO, Fulginiti JV. The influences of student and standardized patient genders on scoring in an objective structured clinical examination. *Acad Med.* 1991;66(9 suppl):S28–S30.
- Carson JA, Peets A, Grant V, McLaughlin K. The effect of gender interactions on students' physical examination ratings in objective structured clinical examination stations. *Acad Med.* 2010;85:1772–1776.
- Haist SA, Wilson JF, Elam CL, Blue AV, Fosson SE. The effect of gender and age on medical school performance: An important interaction. *Adv Health Sci Educ Theory Pract.* 2000;5:197–205.
- Wang-Cheng RM, Fulkerson PK, Barnas GP, Lawrence SL. Effect of student and preceptor gender on clinical grades in an ambulatory care clerkship. *Acad Med.* 1995;70:324–326.
- Rand VE, Hudes ES, Browner WS, Wachter RM, Avins AL. Effect of evaluator and resident gender on the American Board of Internal Medicine evaluation scores. *J Gen Intern Med.* 1998;13:670–674.
- Bienstock JL, Martin S, Tzou W, Fox HE. Medical students' gender is a predictor of success in the obstetrics and gynecology basic clerkship. *Teach Learn Med.* 2002;14:240–243.
- Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of examinee gender and USMLE Step 1 performance. *Acad Med.* 2008;83(10 suppl):S58–S62.
- Cuddy MM, Swanson DB, Clauser BE. A multilevel analysis of the relationships between examinee gender and United States Medical Licensing Exam (USMLE) Step 2 CK content area performance. *Acad Med.* 2007;82(10 suppl):S89–S93.
- Swygert KA, Cuddy MM, van Zanten M, Haist SA, Jobe AC. Gender differences in examinee performance on the Step 2 Clinical Skills data gathering (DG) and patient note (PN) components. *Adv Health Sci Educ Theory Pract.* 2012;17:557–571.
- Austin EJ, Evans P, Goldwater R, Potter V. A preliminary study of emotional intelligence, empathy and exam performance in first year medical students. *Pers Individ Dif.* 2005;39:1395–1405.
- Hojat M, Gonnella JS, Mangione S, et al. Empathy in medical students as related to academic performance, clinical competence and gender. *Med Educ.* 2002;36:522–527.
- Roter DL, Hall JA, Aoki Y. Physician gender effects in medical communication: A meta-analytic review. *JAMA.* 2002;288:756–764.
- Street RL Jr, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician–patient communication to health outcomes. *Patient Educ Couns.* 2009;74:295–301.
- Hall JA, Roter DL. Do patients talk differently to male and female physicians? A meta-analytic review. *Patient Educ Couns.* 2002;48:217–224.
- Hull AL. Medical student performance. A comparison of house officer and attending staff as evaluators. *Eval Health Prof.* 1982;5(1):87–94.
- Spielvogel R, Stednick Z, Beckett L, Latimore D. Sources of variability in medical student evaluations on the internal medicine clinical rotation. *Int J Med Educ.* 2012;3:245–251.