

Emergency Medicine Resident Self-Assessment of Clinical Teaching Compared to Student Evaluation Using a Previously Validated Rubric.

Alan R. Cherney MD, MSc
Lehigh Valley Health Network, Alan_R.Cherney@lvhn.org

Amy B. Smith PhD
Lehigh Valley Health Network, amy_b.smith@lvhn.org

Charles C. Worrilow MD
Lehigh Valley Health Network, charles.worrilow@lvhn.org

Kevin Weaver D.O.,FACOEP
Lehigh Valley Health Network, kevin_r.weaver@lvhn.org

Dawn M. Yenser C-TAGME
Lehigh Valley Health Network, Dawn.Yenser@lvhn.org

See next page for additional authors

Follow this and additional works at: <https://scholarlyworks.lvhn.org/emergency-medicine>



Part of the [Emergency Medicine Commons](#)

Published In/Presented At

Cherney, A. R., Smith, A. B., Worrilow, C. C., Weaver, K. Yenser, D. M., Macfarlan, J. E., Burket, G. A., Koons, A. L., Melder, R. J., Greenberg, M. R., Kane, B. G. (2018). Emergency Medicine Resident Self-Assessment of Clinical Teaching Compared to Student Evaluation Using a Previously Validated Rubric. *Clinical Therapeutics*, 40(8), 1375-1383. doi: 10.1016/j.clinthera.2018.06.013.

This Article is brought to you for free and open access by LVHN Scholarly Works. It has been accepted for inclusion in LVHN Scholarly Works by an authorized administrator. For more information, please contact LibraryServices@lvhn.org.

Authors

Alan R. Cherney MD, MSc; Amy B. Smith PhD; Charles C. WorriLOW MD; Kevin Weaver D.O.,FACOEP; Dawn M. Yenser C-TAGME; Jennifer E. Macfarlan MPH; Glenn A. Burket DO; Andrew L. Koons DO; Raymond J. Melder DO; Marna R. Greenberg DO, MPH, FACEP; and Bryan G. Kane MD



Emergency Medicine Resident Self-assessment of Clinical Teaching Compared to Student Evaluation Using a Previously Validated Rubric

Alan R. Cherney, MD, MSc¹; Amy B. Smith, PhD^{2,3}; Charles C. WorriLOW, MD^{1,3}; Kevin R. Weaver, DO^{1,3}; Dawn Yenser, C-TAGME¹; Jennifer E. Macfarlan, MPH⁴; Glenn A. Burket, DO¹; Andrew L. Koons, DO¹; Raymond J. Melder, DO¹; Marna R. Greenberg, DO, MPH^{1,3}; and Bryan G. Kane, MD^{1,3}

¹Department of Emergency and Hospital Medicine, Lehigh Valley Health Network; ²Department of Education, Lehigh Valley Health Network, Allentown, Pennsylvania; ³Faculty at University of South Florida, Morsani College of Medicine, Tampa, Florida; and ⁴Network Office of Research and Innovation, Lehigh Valley Health Network, Allentown, Pennsylvania

ABSTRACT

Purpose: The quality of clinical teaching in the emergency department from the students' perspective has not been previously described in the literature. Our goals were to assess senior residents' teaching ability from the resident/teacher and student/learner viewpoints for any correlation, and to explore any gender association. The secondary goal was to evaluate the possible impact of gender on the resident/student dyad, an interaction that has previously been studied only in the faculty/student pairing.

Methods: After approval by an institutional review board, a 1-year, grant-funded, single-site, prospective study was implemented at a regional medical campus that sponsors a 4-year dually approved emergency medicine residency. The residency hosts both medical school students (MSs) and physician's assistant students (PAs). Each student and senior resident working concurrently completed a previously validated ER Scale, which measured residents' teaching performance in 4 categories: Didactic, Clinical, Approachable, and Helpful. Students evaluated residents' teaching, while residents self-assessed their performance. The participants' demographic characteristics gathered included prior knowledge of or exposure to clinical teaching models. Gender was self-reported by participants.

The analysis accounted for multiple observations by comparing participants' mean scores.

Findings: Ninety-nine subjects were enrolled; none withdrew consent. Thirty-seven residents (11 women) and 62 students (39 women) from 25 MSs and 6 PA

schools were enrolled, completing 517 teaching assessments. Students evaluated residents more favorably in all ER Scale categories than did residents on self-assessments ($P < 0.0001$). This difference was significant in all subgroup comparisons (types of school versus postgraduate years [PGYs]). Residents' evaluations by type of student (MS vs PA) did not show a significant difference. PGY 3 residents assessed themselves higher in all categories than did PGY 4 residents, with Approachability reaching significance ($P = 0.0105$). Male residents self-assessed their teaching consistently higher than did female residents, significantly so on Clinical ($P = 0.0300$). Students' evaluations of the residents' teaching skills by residents' gender did not reveal gender differences.

Implications: MS and PA students evaluated teaching by EM senior residents statistically significantly higher than did EM residents on self-evaluation when using the ER Scale. Students did not evaluate residents' teaching with any difference by gender, although male residents routinely self-assessed their teaching abilities more positively than did female residents. These findings suggest that, if residency programs utilize resident self-evaluation for programmatic evaluation, the gender of the resident may impact self-scoring. This cohort may inform future study of resident teaching in

Accepted for publication June 18, 2018.

<https://doi.org/10.1016/j.clinthera.2018.06.013>

0149-2918/\$ - see front matter

© 2018 Elsevier Inc. All rights reserved.

the emergency department, such as the design of future resident-as-teacher curricula. (*Clin Ther.* 2018;40:1375–1383) © 2018 Elsevier Inc. All rights reserved.

Key words: clinical teaching, emergency medicine, gender.

INTRODUCTION

The emergency department (ED) presents numerous opportunities for education. Unfortunately, parallel processing of patients, significant patient volume, and inpatient holds are challenges that may impact clinical teaching.^{1–3} As ED shifts rarely permit consistent, dedicated time for educational sessions, the study of teaching in this environment has proven to be difficult.⁴ Most research has examined the dynamic between faculty and residents.² One such prior study investigating the efficacy of teaching by attendings from residents' perspective validated a rubric, known as the ER Scale, to assess clinical teaching in the ED.⁵ The ER Scale converts Likert stems ranging from Unacceptable to Outstanding to a numeric 1 to 5 points for analysis in each of 4 areas: Didactic, Clinical, Approachable, and Helpful.

Despite the challenges of teaching in the ED, previous study has demonstrated that the ED represents a significant part of medical students' (MS) education, notably procedural.⁶ While it has been estimated that ~33% of MSs' learning comes from residents, there is limited literature documenting emergency medicine (EM) residents' teaching effectiveness from the medical student's perspective.⁷ EM residents also supervise other learners in the ED, such as physician's assistant (PA) students. Other specialties with fewer teaching challenges have seen the development of pedagogic models. Examples include the One-Minute Preceptor (or Microskills) and SNAPPS (summarize, narrow, analyze, probe, plan, and select).^{8,9} Both models use a stepwise, systematic approach for case presentations and feedback. This consistent methodology allows learners and teachers to understand expectations and recognize that teaching and learning are to occur.

To support the EM "resident as teacher," a suggested teaching curriculum has been developed.¹⁰ Residents often respond positively when undergoing "resident-as-teacher" curricula, improving their attitudes toward teaching.¹¹ This improved attitude has been demonstrated within EM.¹² There is some evidence that participating in teaching courses correlates

with improved self-assessment¹³ and student evaluations.⁷ Of note, the reviews did not identify a study directly correlating the two.

Despite the limitations of current studies of the efficacy of resident teaching, EM educators are charged with documentation of resident-teaching efficacy. After eliminating a dedicated Resident as Teacher Milestone present in drafts, several of the Accreditation Council for Graduate Medical Education (ACGME) Milestones for EM have, as level 5 anchors, the ability to teach, particularly related to procedures.¹⁴ The suggested evaluation methods for these Milestones focus on simulation, checklists, and procedure logs to assess technical competency.¹⁴ Means of assessing clinical teaching are not formally suggested, which increases the complexity for residency Clinical Competence Committees (CCCs).

Evaluation of teaching through the lens of gender has yielded mixed results. A multispecialty faculty self-assessment and resident evaluation of clinical teaching, which included EM, found gender differences: here women significantly self-assessed their teaching to be better than men.¹⁵ In contrast, when MS rather than residents are the learners, it has been shown that the evaluations of female faculty on the "overall quality of teaching" receive lower scores than male faculty.¹⁶

Given these gaps in the existing literature, the primary aim of our study was to correlate students' evaluations and senior residents' self-assessments of EM residents' clinical teaching on-shift. Secondarily, participant demographic characteristics including gender were analyzed.

MATERIALS AND METHODS

After obtaining our institutional review board's approval, this prospective, observational, educational study was conducted at a suburban health care system serving as a regional allopathic medical campus for dedicated 3rd- and 4th-year MS. It has a 4-year EM residency with both allopathic and osteopathic approval. Within the program, postgraduate year (PGY) 3 and 4 residents are considered senior residents, and supervise students. Our medical school does not have a required 3rd year EM clerkship, and all MSs' EM experience is during a 4th-year elective. The institution also hosts visiting 4th-year MSs from multiple, geographically diverse medical schools, along with PA students from multiple regional schools.

Students are assigned anonymous evaluations of the senior residents with whom they work, electronically via New Innovations software (New Innovations, 3540

Please rate the teaching resident you had contact with in terms of teaching ability.

	Outstanding	Above Average	Average	Below Average	Unacceptable
Didactic	5	4	3	2	1
Clinical	5	4	3	2	1
Approachable	5	4	3	2	1
Helpful	5	4	3	2	1

Figure 1. The ER Scale⁵.

Forest Lake Dr, Uniontown OH 44685. Phone number 330-899-9954), for each shift. Likewise, the senior residents are assigned evaluations of students with whom they work. Either party may decline the evaluation. Due to the ACGME Milestones Project, evaluations were formatted for use of a previously validated 4-question assessment of clinical teaching (ER Scale; Figure), which was chosen for its simplicity and brevity and made standard by the Program Evaluation Committee.⁵

Students were made aware of the both the study and the definitions of the ER Scale scores as part of their rotation orientation. Residents entering PGY 3 or 4 were made aware of the study and provided ER Scale instruction at the annual June residency orientation, and were assigned to read articles on the One-Minute Preceptor for their annual July Journal Club.⁸ Each resident was provided information about the study in writing in case of an absence from orientation.

All students and senior EM residents were eligible to be enrolled in the study. An alteration of informed consent waiver was obtained; residents and students were directed to contact a network educator outside of the residency should they decline to participate. The study was conducted within the EDs of the 2 teaching hospitals in the health care network. The EDs were organized with individual *pods*, or groups consisting of an attending, resident, and student in each. Dedicated teaching shifts occurred 4 times per week and paired a senior resident, junior resident, and student under an attending. During these shifts, the expectations for the senior resident were adjusted. They had assigned teaching tasks, an additional learner (an intern), and decreased patient-throughput responsibilities.

Demographic information collected included age and prior experience with teaching models.^{8,9} Study participants self-reported their gender in a binary fashion. Students reported the number of prior EM rotations, and resident PGY at enrollment was recorded. The study encompassed a single calendar year, from November 15, 2014, to November 14, 2015. One software administrator, the program coordinator, had the ability to retrieve the confidential student evaluations and resident self-assessments. The coordinator did not determine the MS or PA students' clerkship grades or residents' programmatic evaluations, and provided the study team with deidentified data.

Statistical Analysis

Data were analyzed using SAS version 9.3 (SAS Institute, Cary, North Carolina) to compare MS and PA students' ER Scale scores to those of residents. Statistical significance for all comparisons was set at $P \leq 0.05$. Descriptive statistics for students and residents were determined for the 4 categories of the scale: Didactic, Clinical, Approachable, and Helpful.⁵

To allow for independence of observations within the analysis, each participant could have only 1 observation regardless of the number of surveys they filled out. Repeated-measures analysis could not be performed due to unequal survey response rates. Participants (either students or residents) who completed >1 evaluation had the mean score on each ER Scale item calculated.⁵ These mean scores were used for the purposes of hypothesis testing. In those who completed more surveys, data were not adjusted because of potential bias toward improved scores. Residents' scores were analyzed within the PGY of the initial study enrollment.

Table I. Demographics of residents and students. Data are given as number (%) of participants unless otherwise noted.

Characteristic	Residents (n = 37)	Students (n = 62)
Age, mean (SD), y	31.84 (3.15)	28.23 (5.09)
Female	11 (29.7)	32 (51.6)
Medical student	–	45 (72.6)
Physician's assistant	–	17 (27.4)
DO	30 (81.1)	–
Postgraduate year 3	23 (62.2)	–
Postgraduate year 4	14 (37.8)	–
MD	7 (18.9)	–
No prior knowledge of teaching model	20 (54.6)	52 (83.9)

DO = doctor of osteopathy; MD = doctor of medicine.
a: Age is in years; b: All other variables are presented as percentages. The percentages are of volunteered dichotomous information.

Participants were analyzed based on the self-reported (binary) gender. For subgroup analysis, when performing comparisons between residents of different groups (PGY 3 or 4) with student type (MS or PA), students were included in the analysis only if they had completed an evaluation on that specific group of residents.

Wilcoxon scores (rank sums) of each evaluation were calculated for all participants (PGY 3s, PGY 4s, MSs, PAs). Each response was converted to a numeric value for statistical analysis purposes: 1 = unacceptable, 2 = below average, 3 = average, 4 = above average, 5 = and outstanding. Steiner et al⁵ described this conversion to a 5-point Likert scale. A Wilcoxon 2-sample test was performed and 2-sided *P* values were calculated.

RESULTS

Over 12 months, 99 subjects were enrolled (Table I). None of the residents or students contacted the network educator to opt out of the study. Of the 37 residents eligible, all of whom enrolled, 11 (29.7%) were women. Of the 62 students, 39 (51.6%) were women. The 62 students represented 25 medical and 6 distinct PA schools. Twenty students (32.3%) participated during the 2014–2015 academic year, with the remaining 42 participating in 2015–2016. Students' prior knowledge of teaching models was limited; 83.9% of students had

no prior knowledge of either model. Thirty-nine of the students (62.9%) indicated that they were pursuing a career in EM. The experience with previous EM rotations varied; the majority (88.1%) stated they were in the 1st through 4th rotation. The prior duration of clinical teaching education most commonly reported by residents was 1 to 4 hours (31%), with 76.8% having ≥ 1 hour of training. Only 38.1% of residents reported exposure to the teaching models.

A total of 517 on-shift teaching assessments from students and residents were analyzed. The majority of students rated the residents as either outstanding or above average on all 4 items of the ER Scale (Table II).⁵ This finding is in contrast to residents, who generally marked themselves as either average or above average (Table III). Approximately one third of residents and students indicated either “not enough time” or simply did not complete the evaluation.

There was a statistically significant difference between resident self-assessment and student evaluation scores on the ER Scale (Table IV), with students' evaluation scores being higher than residents' self-assessment scores.⁵ This difference was statistically significant on all ER Scale items when comparing residents to MS and again to PA students. When subgroup analysis comparing PGY 3 residents to all students, as well as the subgroups of MS and PA students was performed, statistical significance also was found on all ER Scale items.⁵ The differences in each case noted *P* values ≤ 0.0001 , except for PGY 3 versus PA in the category of Approachable (*P* = 0.0017). This same analysis in PGY 4 residents found statistically significant differences on all ER Scale items. Comparisons of PGY 4 versus all students, as well as the subgroups of MS and PA students had *P* values ≤ 0.0001 on all ER Scale items.⁵ Students' evaluations of residents did not differ with significance when comparing MS to PA students (Didactic, *P* = 0.3044; Clinical, *P* = 0.2469; Approachable, *P* = 0.1064; Helpful, *P* = 0.4333).⁵

Residents were compared to each other to determine whether there was a difference between PGY 3 and PGY 4 self-evaluations. The Approachable item showed significance (*P* = 0.0105), whereas the Helpful item approached significance (*P* = 0.0505), with PGY 3 residents' self-evaluation scores on all categories being higher.⁵ Residents and students were evaluated by shift type. Students scheduled on dedicated teaching shifts submitted ratings similar to those of nonteaching shifts.

Aggregate student assessment of the residents' teaching skills by resident gender did not reveal differences:

Table II. Student assessment of resident teaching using the ER scale. Data are given as number (%) of participants.

Student Type	Didactic	Clinical	Approachable	Helpful
All students (n = 62)				
Outstanding	180 (34.8)	196 (37.9)	275 (53.2)	297 (57.4)
Above average	144 (27.8)	137 (26.5)	70 (13.5)	53 (10.3)
Average	32 (6.2)	25 (4.8)	13 (2.5)	9 (1.7)
Below average	1 (0.2)	2 (0.4)	2 (0.4)	1 (0.2)
Unacceptable	0	0	0	0
Not enough time	34 (6.6)	34 (6.6)	34 (6.6)	34 (6.6)
Missing	126 (24.4)	123 (23.8)	123 (23.8)	123 (23.8)
Medical students (n = 45)				
Outstanding	113 (32.3)	126 (36.0)	196 (56.0)	192 (54.8)
Above average	102 (29.1)	97 (27.7)	41 (11.7)	45 (12.9)
Average	28 (8.0)	22 (6.3)	8 (2.3)	9 (2.6)
Below average	1 (0.3)	2 (0.6)	2 (0.6)	1 (0.3)
Unacceptable	0	0	0	0
Not enough time	19 (5.4)	19 (5.4)	19 (5.4)	19 (5.4)
Missing	87 (24.9)	84 (24.0)	84 (24.0)	84 (24.0)
PA students (n = 17)				
Outstanding	67 (40.1)	70 (41.9)	79 (47.3)	105 (62.9)
Above average	42 (25.2)	40 (24.0)	29 (17.4)	8 (4.8)
Average	4 (2.4)	3 (1.8)	5 (3.0)	0
Below average	0	0	0	0
Unacceptable	0	0	0	0
Not enough time	15 (9.0)	15 (9.0)	15 (9.0)	15 (9.0)
Missing	39 (23.3)	39 (23.3)	39 (23.3)	39 (23.3)

PA = physician's assistant.

(Didactic: men, 4.50 vs women, 4.18 [$P = 0.1427$]; Clinical: men, 4.50 vs women, 4.36 [$P = 0.4647$]; Approachable: men, 4.69 vs women, 4.73 [$P = 0.8502$]; Helpful: men, 4.96 vs women, 4.82 [$P = 0.1600$]).⁵ By contrast, as shown in Table V, the residents' self-assessments did show differences by gender. In all cases, the men rated themselves more highly than did the women, with Clinical reaching the level of statistical significance ($P = 0.0300$). Of note, none of the female residents rated themselves as outstanding on any of the surveys they completed, while men did this several times.

DISCUSSION

This study sought first to determine the correlation between students' assessments and residents' self-evaluations of clinical teaching in the ED. The study findings

are discussed in that order: students' assessments and residents' self-evaluations.

Our study found that students—both MS and PA—appear to evaluate senior EM residents' teaching in the ED significantly more favorably than residents assess themselves. Nearly two thirds of the student evaluations rated the residents as above average or outstanding in each of the 4 ER Scale categories.⁵ This finding suggests that students positively perceive EM residents' teaching. That this positive perception was developed from students from a large cross section of schools likely increases the generalizability of the findings from the cohort. Similar findings of students' perception of EM residents' clinical teaching have not been previously reported. One prior study in a simulation laboratory found that EM residents' teaching as assessed by students was similar to

Table III. Resident self-assessment of their teaching using the ER scale.

Resident Type	Didactic	Clinical	Approachable	Helpful
All residents (n = 37)				
Outstanding	25 (4.8)	25 (4.8)	48 (9.3)	24 (4.6)
Above average	159 (30.8)	181 (35.0)	175 (33.8)	183 (35.4)
Average	147 (28.4)	125 (24.2)	112 (21.7)	126 (24.4)
Below average	6 (1.2)	5 (1.0)	2 (0.4)	3 (0.6)
Unacceptable	0	0	0	0
Not enough time	129 (24.9)	129 (24.9)	129 (24.9)	129 (24.9)
Missing	51 (9.9)	52 (10.1)	51 (9.9)	52 (10.1)
PGY 3 (n = 23)				
Outstanding	2 (1.0)	4 (2.0)	18 (8.9)	5 (2.5)
Above average	65 (32.3)	82 (40.8)	88 (43.8)	87 (43.3)
Average	83 (41.3)	65 (32.3)	49 (24.4)	60 (29.8)
Below average	5 (2.5)	4 (2.0)	0	3 (1.5)
Unacceptable	0	0	0	0
Not enough time	34 (16.9)	34 (16.9)	34 (16.9)	34 (16.9)
Missing	12 (6.0)	12 (6.0)	12 (6.0)	12 (6.0)
PGY 4 (n = 14)				
Outstanding	23 (7.3)	21 (6.6)	30 (9.5)	19 (6.0)
Above average	94 (29.8)	99 (31.3)	87 (27.5)	96 (30.3)
Average	64 (20.2)	60 (19.0)	63 (20.0)	66 (20.9)
Below average	1 (0.3)	1 (0.3)	2 (0.6)	0
Unacceptable	0	0	0	0
Not enough time	95 (30.1)	95 (30.1)	95 (30.1)	95 (30.1)
Missing	39 (12.3)	40 (12.7)	39 (12.3)	40 (12.7)

PGY = postgraduate year.

Table IV. Comparison of mean (SD) ER scale scores between residents and students.

Domain	Residents (n = 37)	Students (n = 62)	P
Didactic	3.46	4.40	<0.0001
Clinical	3.49	4.44	<0.0001
Approachable	3.76	4.68	<0.0001
Helpful	3.65	4.73	<0.0001

Scale: unacceptable = 1; below average = 2; average = 3; above average = 4; outstanding = 5.

that of attendings.¹⁷ The findings from that study appear to support our conclusion that in the ED, both residents and attendings may effectively teach students.

It is notable that, in all comparisons, residents in this cohort scored themselves statistically significantly

lower than did students on evaluations. Self-assessment is a behavioral anchor as part of the ACGME Milestone Project, so an understanding of that process by EM residents informs in part how EM educators should utilize that information.¹⁴ A systematic review found that

Table V. Gender comparisons of mean scale scores on residents' self-evaluations.

Category	Male (n = 26)	Female (n = 11)	<i>P</i>
Didactic	3.58	3.18	0.0707
Clinical	3.62	3.18	0.0300
Approachable	3.85	3.55	0.2267
Helpful	3.77	3.36	0.0568

Scale: unacceptable = 1; below average = 2; average = 3; above average = 4; outstanding = 5.

physicians' self-assessments—when compared to objective measures—were limited and had a wide range of variability.¹⁸ Within EM, the accuracy of residents' self-assessment in the simulation laboratory appears to be tied to performance, with higher performers being more accurate.¹⁹ The literature notes that difficulty calibrating a clinical teaching gold standard can significantly affect self-assessment scores, as can viewpoints differing between faculty and learners.²⁰ Therefore, a definition of clinical teaching should be made with specific tools for assessment. This definition will improve Milestone reporting accuracy, potentially allowing for residents' self-assessment of their teaching. In our cohort, it is possible the PGY 3s self-assessed their teaching more favorably on the ER Scale than did PGY 4s because the more senior residents realized how difficult teaching can be. Despite the limitations of self-assessment, the reflective process, especially when combined with student feedback, can be a strong impetus to improve.²¹

As a secondary measure, the study sought to evaluate teaching through the lens of gender. This secondary analysis is discussed by focus on the students' assessments first, followed by residents' self-assessments. This cohort found no significant difference in students' evaluation of residents' teaching by gender. This finding is contrast to those from the publication by Morgan et al,¹⁶ which found significantly lower students' evaluation scores from MSs of female faculty as compared to male clinical faculty. That our cohort showed no differences in the scores on the evaluations of clinical teaching by students in a variety of resident-teacher/student-learner dyads suggests that the scheduling of house staff and learners may be done independent of one another without affecting the learning environment of students. It should be noted

that Morgan's study enrolled students from within a single medical school, whereas our cohort was composed primarily of visiting students, which may explain the different findings.¹⁶ The majority of other published analyses of gender comparisons of evaluations of teaching have come from resident learners with some studies finding that male clinical teachers receive better evaluations, others finding that female faculty receive higher ratings in specific areas, and finally some of which finding no relationship.¹⁵ Therefore, another explanation of the findings that in the present cohort the gender of the teacher and that of the learner did not appear to affect the education that the students received may be that students view clinical teachers differently than resident learners do.

The scores on self-assessments of female residents in the present cohort were universally less than those of their male colleagues. This finding is in contrast to a cohort in which female faculty assessed their teaching ability significantly higher than did male faculty.¹⁵ That study used a summative instrument rather than a formative one such as the ER scale, which required upward of 10 minutes to complete.²² Perhaps the immediacy of the assessment in the present cohort, with an instrument that required seconds rather than minutes to complete, contributed to the difference. The educational literature notes that the gender connotations of evaluation stem words may affect the self-assessment ratings of male versus female teachers.²³ Given the single-word stems in the ER Scale, this effect may have played a role here, and, if so, validated gender-neutral rubrics will need to be developed.⁵

Finally, reported awareness of specific teaching models in our cohort is low, suggesting an opportunity for improvement. Awareness of these models and rubrics is important for participation and successful self-assessment. Neither the teaching models mentioned previously (One Minute Clinical Preceptor and SNAPPS) nor the impact of "resident as teacher" curricula on medical student evaluations has been evaluated through the lens of gender.^{8,9} Women in medicine, in accordance with adult-learning theory, may feel the need to prove themselves constantly;²⁴ thus, it may be necessary to develop gender-specific pedagogic models or resident-as-teacher curricula to help women residents feel effective as clinical teachers. In this cohort, women were statistically more likely than men to underrate themselves. Research from numerous corners of psychological inquiry suggest that self-assessments of skill are often flawed; the correlation between self-ratings of skill performance and

actual performance is often moderate to meager because people overrate themselves.²⁵ Having gender-specific courses may imbue female teachers with more confidence. Moving forward, the findings from this cohort suggest that evaluation of residents' teaching skills by students is feasible. The findings from the present study may serve as a baseline for future investigation into the resident–student teaching relationship within EM, including that of interventions aimed at supporting residents in their teaching role. Given the gender differences in residents' self-evaluation scores noted in this cohort, effective interventions may differ by gender. Programs seeking to utilize self-evaluation as part of the summative residency evaluation process may need to consider the gender of the residents before finalizing the Milestone scores in the Clinical Competence Committee.

Limitations

This study was limited by having been performed at a single site, where the nature of the resident–student dyad is typically one on one. This focused attention may have positively influenced the students' evaluations. That said, the findings did not differ significantly by teaching shift when multiple learners were present. The residents were aware that the students would evaluate their teaching, which may have created a Hawthorne effect, further biasing the students' evaluations positively. Students' grading concerns may have influenced the outcomes, likely in a positive direction. Approximately one third of shift evaluations were not completed, which may have further biased the results, likely overstating the quality of the teaching. Given the unequal participation of the participants, hypothesis testing was performed using the mean score of each enrolled subject. By using means, the scores of infrequent (especially single submission) participants had more influence on the results than did those of more frequent participants. A power calculation was not performed; rather, the design was over 365 days spanning 2 academic years. The analytic plan accounted for this academic transition of enrollees by conducting residents' analysis in the year enrolled, such that by the end of the study, an enrolled PGY 3 may have been entering ER Scale data while working as a PGY 4.

CONCLUSIONS

In the present cohort, MS and PA students evaluated teaching by EM senior residents statistically significantly higher than did EM residents on self-evaluation

using the ER Scale. This significant difference was found in all subgroup comparisons by type of student and PGY year. Students had a positive view of residents' teaching in the ED and did not evaluate quality of teaching by residents differently by gender. These findings may have implications for staffing of students with teachers, including scheduling. Male EM residents consistently self-assessed themselves higher than female EM residents, sometimes with statistical significance. This difference should be considered if residency programs seek to use self-evaluation as part of residents' summative evaluations.

ACKNOWLEDGMENTS

The authors acknowledge Hope M. Kincaid, MPH, CPH, for statistical support; Kristine A. Petre, MLS, CM, AHIP, for her assistance with obtaining the literature; Anita Kurt, PhD, RN, for research operations and management; and Bernadette Glenn-Porter, BS, for editing and formatting this article. Funding, in part, for this study was received from the Anne and Carl Anderson Trust, a nonprofit community trust foundation.

All of the authors made substantial contributions to this work and approved the final manuscript. ARC, Literature search, protocol development, study execution, data gathering, analysis, table development, writing, editing; ABS, Protocol development, consent determination, editing; CCW, Protocol development, study execution, edits; KRW, Protocol development, study execution, edits; DY, Protocol development, study execution, data gathering, edits; JEM, Protocol development, analysis, tables, writing, edits; GAB, Protocol development, study execution, edits; ALK, Protocol development, study execution, edits; RJM, Protocol development, study execution, edits; MRG, Analysis, tables, writing, edits; BGK, Literature search, protocol development, study execution, data gathering, analysis, tables, writing, edits.

CONFLICTS OF INTEREST

The Anderson Trust funding was an unrestricted educational grant. The authors have no other financial disclosures for this study.

REFERENCES

1. Aldeen AZ, Gisondi MA. Bedside teaching in the emergency department. *Acad Emerg Med*. 2006;13:860–866.

2. Atzema C, Bandiera G, Schull MJ. Emergency department crowding: the effect on resident education. *Ann Emerg Med.* 2005;45:276–281.
3. Pines JM, Prabhu A, McCusker CM, Hollander JE. The effect of ED crowding on education. *Am J Emerg Med.* 2010;28:217–220.
4. Ilgen JS, Takayesu JK, Bhatia K, et al. Back to the bedside: the 8-year evaluation of a resident-as-teacher rotation. *J Emerg Med.* 2011;41:190–195.
5. Steiner IP, Franc-Law J, Kelly KD, Rowe BH. Faculty evaluation by residents in an emergency medicine program: a new evaluation instrument. *Acad Emerg Med.* 2000;7:1015–1021.
6. Johnson GA, Pipas L, Newman-Palmer NB, Brown LH. The emergency medicine rotation: a unique experience for medical students. *J Emerg Med.* 2002;22:307–311.
7. Post RE, Quattlebaum RG, Benich 3rd JJ. Residents-as-teachers curricula: a critical review. *Acad Med.* 2009;84:374–380.
8. Neher JO, Gordon KC, Meyer B, Stevens N. A five-step "microskills" model of clinical teaching. *J Am Board Fam Pract.* 1992;5:419–424.
9. Wolpaw T, Papp KK, Bordage G. Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: a randomized comparison group trial. *Acad Med.* 2009;84:517–524.
10. Farrell SE, Pacella C, Egan D, et al. Society for Academic Emergency Medicine, Undergraduate Education Committee. Resident-as-teacher: a suggested curriculum for emergency medicine. *Acad Emerg Med.* 2006;13:677–679.
11. Wipf JE, Pinsky LE, Burke W. Turning interns into senior residents: Preparing residents for their teaching and leadership roles. *Acad Med.* 1995;70:591–596.
12. Wachtel JK, Greenberg MR, Smith AB, Weaver KR, Kane BG. Residents as teachers: residents' perceptions before and after receiving instruction in clinical teaching. *J Am Osteopath Assoc.* 2013;113:23–33.
13. Wamsley MA, Julian KA, Wipf JE. A literature review of "resident-as-teacher" curricula: do teaching courses make a difference? *J Gen Intern Med.* 2004;19(5 Pt 2):574–581.
14. The Emergency Medicine Milestones Project. The Accreditation Council for Graduate Medical Education; 2018. Accessed June 11; 2018. <https://www.acgme.org/acgmeweb/Portals/0/PDFs/Milestones/EmergencyMedicineMilestones.pdf>. Accessed June 11.
15. Fluit CR, Feskens R, Bolhuis S, Grol R, Wensing M, Laan R. Understanding resident ratings of teaching in the workplace: a multi-centre study. *Adv Health Sci Educ Theory Pract.* 2015;20:691–707.
16. Morgan HK, Purkiss JA, Porter AC, Lyson ML, Santen SA, Christner JG, Grum CM, Hammoud MM. Student evaluation of faculty physicians: gender differences in teaching evaluations. *J Womens Health (Larchmt).* 2016;25:453–456.
17. Cooper DD, Wilson AB, Huffman GN, Humbert AJ. Medical students' perception of residents as teachers: comparing effectiveness of residents and faculty during simulation debriefings. *J Grad Med Educ.* 2012;4:486–489.
18. Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA.* 2006;296:1094–1102.
19. Sadosty AT, Bellolio MF, Laack TA, Luke A, Weaver A, Goyal DG. Simulation-based emergency medicine resident self-assessment. *J Emerg Med.* 2001;41:679–685.
20. Ward M, Gruppen L, Regehr G. Measuring self-assessment: current state of the art. *Adv Health Sci Educ Theory Pract.* 2002;7:63–80.
21. Stalmeijer RE, Dolmans DH, Wolfhagen IH, Peters WG, van Coppenolle L, Scherpbier AJ. Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Adv Health Sci Educ Theory Pract.* 2010;15:315–328.
22. Fluit C, Bolhuis S, Grol R, Ham M, Feskens R, Laan R, et al. Evaluation and feedback for effective clinical teaching in postgraduate medical education: validation of an assessment instrument incorporating the CanMEDS roles. *Med Teacher.* 2012;34:893–901.
23. Masunaga H, Hitchcock MA. How does gender interact with clinical teachers' perceptions of clinical teaching? *NASPA J Women Higher Educ.* 2011;4:50–72.
24. Knowles M. *The Adult Learner: A Neglected Species.* Houston, TX: Gulf Publishing; 1990.
25. Dunning D, Heath C, Suls J. Flawed self-assessment—implications for health, education and the workplace. *Psychol Sci Public Interest.* 2004;5:69–106.

Address correspondence to: Bryan G. Kane, MD, Emergency Medicine Residency Program, 5th Floor, EM Residency Suite, 2545 Schoenersville Road, Bethlehem, PA 18017. E-mail: Bryan.Kane@lvhn.org