

New directions in medical student clerkship evaluations

Avery M Whitis, MD,¹ Abbey Hardy-Fairbanks, MD,¹ Colleen K Stockdale, MD, MS¹

Keywords: Evaluation; medical student; grade

Abstract

Purpose: To compare the number of requested medical student evaluations with mean evaluation scores and final clerkship grades.

Background: The University of Iowa Obstetrics & Gynecology (OBGYN) clerkship requires six evaluations for each student: two assigned and four requested by the student. Many students request more evaluations than required, contributing to a backlog that slows the grading process.

Methods: Medical student evaluations from 2014 to 2016 were analyzed. Three groups were created based on the number of evaluations an individual student received. Group 1 received 3-4 evaluations (73), Group 2 received 5-6 evaluations (240) and Group 3 received 7-16 evaluations (222). A paired T-test compared mean evaluation scores and a chi-square test was used to compare mean shelf exam scores and percentages of pass, near honors, and honors grades.

Results: A total of 535 independent students and their evaluations were reviewed for the study. The difference in mean evaluation scores for groups 1 and 2 and groups 2 and 3 were statistically significant ($p < 0.01$, $p = 0.02$, respectively). The differences in mean shelf grades between groups 1 and 2 and groups 2

and 3 were not significant ($p > 0.05$). Similarly, the differences in rates of pass, near honors, and honors grades between groups were not statistically significant ($p > 0.05$).

Discussion: Increased number of requested evaluations did not translate to differences in rates of pass, near honors, and honors grades for medical students on their OBGYN clerkship, indicating that decreasing the mandatory evaluations per student would not be detrimental to student outcomes and would potentially expedite the grading process.

¹Department of Obstetrics and Gynecology, University of Iowa Carver College of Medicine, Iowa City, Iowa

Introduction

The Accreditation Council for Graduate Medical Education (ACGME) provides six core competencies to guide medical student learning: practice based learning and improvement, patient care and procedural skills, systems based practice, medical knowledge, interpersonal and communication skills, and professionalism.¹ These core competencies are assessed through

Please cite this paper as: Whitis AM, Hardy-Fairbanks A, Stockdale CK. New directions in medical student clerkship evaluations. POG in Press. 2019 November; Article 3 [5 p.]. Available from: http://ir.uiowa.edu/pog_in_press/. Free full text article.

Corresponding author: Avery M Whitis, MD, Department of Obstetrics and Gynecology, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, IA 52246; 319-356-2291, avery-whitis@uiowa.edu

Financial Disclosure: The authors report no conflict of interest.

Copyright: © 2019 Whitis et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

written exams, simulated patient experiences, written notes, and formal evaluations. Evaluations are a valuable way to assess nontechnical skills (e.g. professionalism, empathy, team interaction), are low cost and can be tailored from institution to institution. However, they have been shown to correlate poorly with fund of knowledge exams² and are subject to bias and error.^{3,4} Often there are time delays between initial teacher/learner exposure and evaluation. Furthermore, faculty in a program can have widely different teaching and evaluation expertise.

At the University of Iowa, grading for the mandatory Obstetrics & Gynecology (OBGYN) clerkship is broken down into four components: written exam (45%), assignments (15%), participation (20%), and evaluations (20%). A minimum of six evaluations were required, two from each of the main components of the clerkship (inpatient gynecology, inpatient obstetrics, and clinic). Four of these evaluations were requested by the student to the evaluator of their choice, and two were assigned by the program. However, students could request as many evaluations as they chose, and many would request more than the six required, contributing to a backlog of evaluations that slowed the grading process. To optimize timely evaluation and reduce evaluator burden the evaluation process was restructured to include assigned team evaluations. Students are now required to have a total of four evaluations. This facilitated a transition from individual evaluations by a single teacher to team evaluations with feedback pooled from the resident, fellow, and attending teams that interacted with the student.

This study was undertaken to compare mean evaluation scores, shelf exam scores, and final clerkship grades to determine if the change in evaluation process would adversely affect individual student grades.

Methods

Medical student evaluations from 2014 to 2016 were analyzed. Three separate groups of students were created based on the number of evaluations an individual student received. Group 1 received 3-4 evaluations (n=73), group 2 received 5-6 evaluations (n=240), and group 3 received 7-16 evaluations (n=222). A paired T-test was used to compare mean evaluation scores between the groups using group 2 as the reference (6 evaluations historically required) to compare both fewer and greater number of individual evaluations. Chi-square test was used to compare mean shelf exam scores and percentages of pass, near honors, and honors grades between the groups. Statistical analysis completed with OpenEpi[®], open access epidemiologic software, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2663701/>.⁵

Results

A total of 535 students participated in the study. Mean evaluation scores were 86.75, 89.08, and 88.70 for groups 1, 2, and 3, respectively. The difference in mean evaluation scores for groups 1 and 2 and groups 2 and 3 were statistically significant (p < 0.01, p = 0.02, respectively). Mean shelf exam scores were 77.19, 77.79, and 78.11 for groups 1, 2, and 3, respectively. The differences in mean shelf exam scores

were not significant ($p = 0.53$, $p = 0.35$, for groups 1 and 2, and groups 2 and 3, respectively).

groups were not statistically significant (pass: $p = 0.69$, 0.44 ; near honors: $p = 0.47$, 0.95 ; honors: $p = 0.68$, 0.32 for groups 1 and 2, and groups 2 and 3, respectively).

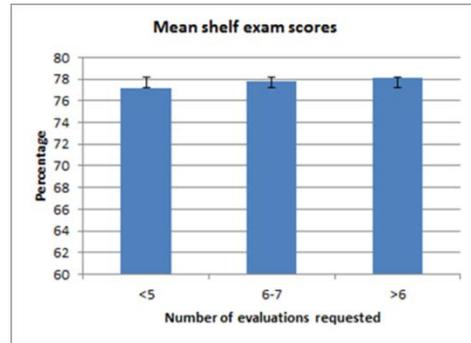
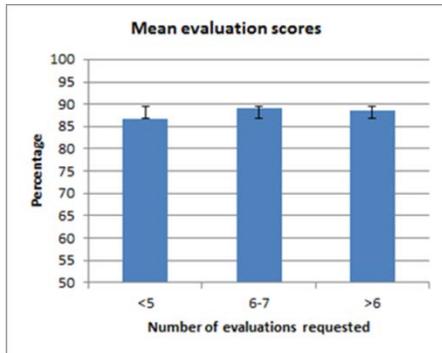


Figure 1. Difference in mean evaluation scores between groups

Figure 2. Mean shelf exam scores between groups. All p values > 0.05

Similarly, the difference in the percentage of pass, near honors, and honors clerkship grades between

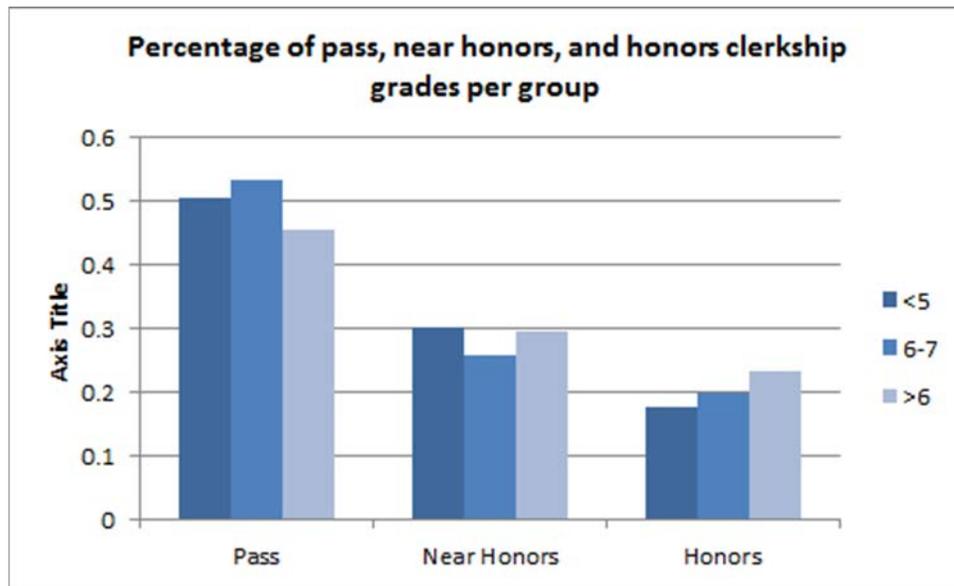


Figure 3. Final clerkship grades among study groups. All p values > 0.05

Discussion

While an increased number of requested evaluations was correlated with higher mean total scores on the evaluation, an increased number of requested evaluations did not translate to differences in rates of pass, near honors, and honors grades for medical students on their OBGYN clerkship. This supports the theory that decreasing the mandatory evaluation number per student would not be detrimental to student outcomes and would potentially expedite the grading process. Additionally, increasing the number of teachers involved in each evaluation process (i.e. expanding on team evaluations) should aid in accuracy of evaluation scores. This idea has previously been discussed - Battinstone, et al., demonstrated that increasing the use of group evaluations may decrease bias, inflation, and improve evaluator competency.⁶ Additionally, team input functionally increases the weight of each individual score and may be in alignment with prior estimates that 7-11 evaluations over a broad range of encounters are needed for accurate student assessment.⁷

Limitations to this study include group distribution as group 1 was significantly smaller than the other groups. Our results are also limited by small sample size as well as restriction to one medical education center.

In conclusion, our data demonstrate that increasing total number of student evaluations does not translate to higher grades in the Obstetrics and Gynecology clerkship at the University

of Iowa. Our institution hopes to use this data to continue to improve the quality and timeliness of the grading and evaluation process for our students during the clerkship as well as inform evaluation processes throughout the institution as a whole.

References

1. Milestones. Chicago, IL: Accreditation Council for Graduate Medical Education(ACGME), c2000-2019. <https://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview>
2. Dubosh NM, Fisher J, Lewis J, Ullman EA. Faculty Evaluations Correlate Poorly with Medical Student Examination Performance in a Fourth-Year Emergency Medicine Clerkship. *J Emerg Med.* 2017 Jun;52(6):850-855. <https://doi.org/10.1016/j.jemermed.2016.09.018> Epub 2017 Mar 22. PubMed PMID: 28341085.
3. Bowen RE, Grant WJ, Schenarts KD. The sum is greater than its parts: clinical evaluations and grade inflation in the surgery clerkship. *Am J Surg.* 2015 Apr;209(4):760-4. <https://doi.org/10.1016/j.amjsurg.2014.10.023> Epub 2015 Jan 22. PubMed PMID: 25707366.
4. Reid CM, Kim DY, Mandel J, Smith A, Bansal V. Correlating surgical clerkship evaluations with performance on the National Board of Medical Examiners examination. *J Surg Res.* 2014 Jul;190(1):29-35. <https://doi.org/10.1016/j.jss.2014.02.031> Epub 2014 Feb 23. PubMed PMID: 24666987.
5. Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version. www.OpenEpi.com updated 2013/04/06, accessed 2019/11/18.

6. Battistone MJ, Pendleton B, Milne C, Battistone ML, Sande MA, Hemmer PA, Shomaker TS. Global descriptive evaluations are more responsive than global numeric ratings in detecting students' progress during the inpatient portion of an internal medicine clerkship. *Acad Med.* 2001 Oct;76(10 Suppl):S105-7.
<https://doi.org/10.1097/00001888-200110001-00035> PubMed PMID: 11597889.
7. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med.* 2003 Fall;15(4):270-92.
https://doi.org/10.1207/S15328015TLM1504_11 PubMed PMID: 14612262.