

Evaluating Evaluation: What, How, and Why?

Joshua Freeman, MD

(Fam Med. 2019;51(10):801-2.)

doi: 10.22454/FamMed.2019.344744

To understand the outcomes of medical education programs—that is, the effectiveness of medical schools—requires multiple measures. We traditionally assess knowledge and skill, using exams and faculty evaluations. Some programs use standardized patients to evaluate specific clinical skills. But in order to comprehensively understand the impact of our educational programs, we also need to look at specialty choice, practice location, and the populations cared for by graduates. If those graduates do not apply the knowledge and skills they have learned to benefit certain populations based, for example, upon geography, socioeconomic status, or ethnicity, then their overall impact on the public's health might be compromised. When graduates choose specialties and practice locations where there is no shortage, and ignore specialties and practice locations where they are needed, they increase the degree to which our health care system is inequitable, rather than helping to ameliorate it.

If we want the outcomes of medical education to be different, we also need to examine how we assess and evaluate our students. If the criteria we use to evaluate them is mainly retention of information, we might not be preparing them to practice in the communities they will eventually serve. Facts are easily accessible in our digital world, but students must learn how to integrate and use them to effectively help people to be healthier. More importantly, many students who become excellent clinicians do not perform well on standardized content-based examinations, so placing too much emphasis on such exams might end up creating a system that de facto discriminates against such students.

This issue of *Family Medicine* includes two articles that address the teaching and

evaluation of students receiving clinical training in our discipline. This training occurs on our family medicine clerkships, usually by placing students with practicing family physicians in the community. Drowos and colleagues address the challenges of finding sufficient numbers of preceptor sites for family medicine clerks and ensuring that students receive high-quality teaching in them.¹ Their two important findings were: (1) schools that pay preceptors (or their employers) have less difficulty finding them in the first place and an easier time getting them to accept remediation when their teaching is assessed as deficient, and (2) departments that offer remediation for preceptors are less likely to have to eliminate underperforming sites. The first finding recognizes that preceptors' time is valuable, while the second validates a key principle of adult learning: improvement is most likely when structured opportunities for remediation are available.

The study by Drs Zoberi Schiel and Everard asked family medicine clerkship directors in a CERA survey if they considered grade inflation to be a problem, and if so, what they are doing about it and whether it is working.² The authors found that a majority believed grade inflation is a problem, and that those in this majority are more likely to use objective criteria for evaluation. They are also more likely to have a structure for providing feedback to routinely high graders. Moving away from subjective clinical evaluations can, however, create a conundrum. On the one hand, when used well, subjective evaluations might be more accurate assessments of clinical performance than knowledge-based tests. On the negative side, subjective evaluations can potentially both inflate grades (if the grader is inexperienced or conflict-averse) and discriminate against certain groups of students.

To be sure, discrimination in medical student evaluations remains a concern. In a recent essay in the *New England Journal of Medicine*,³ Derek J. Paul, a UCSF medical student, begins with a discussion of the photos in the medical school yearbook of Virginia Governor Ralph Northam that depict students dressed in blackface and in Klan robes. He notes that

...after 20 years of diversity- and equity-training efforts, and although we have in fact diversified our medical school classes, racial minorities remain grossly underrepresented among faculty in academic medicine.

This is a truth that should embarrass us all. He observes that he is "...part of the first class for which the UCSF School of Medicine has shifted from a grading system of honors/pass/fail in the core clerkship to a pass/fail system," a change that "came after UCSF openly studied and published grading data that showed persistent disparities affecting student groups that are underrepresented in medicine." But Paul also raises concern that this will

...put more emphasis on qualitative evaluations, which have been shown in a national study to result in disparities of a different type: white students are statistically more likely to be described as 'exceptional,' black students as 'competent,' and female students as 'caring' or 'empathic.'

Discrimination against both individuals and groups from nonmajority backgrounds, intentional or not, is completely unacceptable and a violation of basic human rights, but the reality is that unconscious bias exists and must be addressed. If we truly want to address disparities in health care, we must train more minority physicians, who are more likely to practice with populations and in areas of physician shortage.⁴ We also need minority physicians to participate in far greater numbers in the process of teaching the next generation. If the evidence shows that students from underrepresented minority groups do less well on written tests, or get lower evaluations by subjective clinical graders, we must make a priority of fixing our flawed evaluation systems. This will require changes in who we use as preceptors, how those preceptors complete the evaluation process, and how evaluations are used to help students reach their potential. We must be able to get good and accurate

assessments of a student's clinical skill while guarding against discrimination arising from both unconscious and conscious bias.

The best way to do this is not yet completely clear, but it is likely to include pass/fail grading, objective structured clinical examinations, and qualitative clinical evaluations that are guided by structured objective criteria. The latter may be less likely to be skewed by style or personality differences between evaluators and students. Indeed, it may well be that it is exactly the way in which those students are different that leads them to have a more positive interaction with some patients who have difficulty with doctors like those performing the evaluations! The higher level of faculty skill that will be needed means that we will need more rigorous faculty development activities than we currently have in place, as suggested by Drowos et al. However, by most evidence, we have fewer such opportunities than in the past. Faculty development will have to be funded, and it will have to be a priority. If we are to achieve greater equity in health care we need to produce physicians who will care for all Americans, and indeed, who look like the American population. Yes, they must be skilled and knowledgeable, but the methods we use to measure them must be designed to ensure that they do not obstruct this essential goal.

While our current evaluation system is probably more negative for underrepresented minority students, it really serves none of our students well. We seek to measure competency rather than numbers, but there is little proof that our current faculty—particularly preceptors—know how to do this well. We must have preceptors whose student evaluations are both reliable and valid, and not accept those who are not.

CORRESPONDENCE: Address correspondence to Dr Joshua Freeman at jfreeman@kumc.edu.

References

1. Drowos J, Sairenji T, Hood Watson K, et al., Identifying and remediating quality issues at clinical teaching sites: a CERA Clerkship Directors Survey. *Fam Med*. 2019;51(10):811-816.
2. Zoberi Schiel K, Everard KM, Grade inflation in the family medicine clerkship. *Fam Med*. 2019;51(10):806-810.
3. Paul DW Jr. Ghosts of our collective subconscious - what blackface in a yearbook photo means for medical education. *N Engl J Med*. 2019;381(5):402-403.
4. Komaromy M, Grumbach K, Drake M, et al. The role of black and Hispanic physicians in providing health care for underserved populations. *N Engl J Med*. 1996;334(20):1305-1310.