Virtual Mentor

CLINICAL CASE
Can a Pass/Fail Grading System Adequately Reflect Student Progress?
Commentary by Bonnie M. Miller, MD, Adina Kalet, MD, MPH, Ryan C. VanWoerkom, Nicholas Zorko, and Julia Halsey

As David, a second-year medical student, made his way into the lecture hall, he was surprised to see how packed the room was. A group of 25 third-year students, or one-fifth of the class, had recently petitioned to switch from a traditional letter-grade system to one that was pass/fail at their school, and the medical student government organized a townhall meeting for students to discuss the matter. Unable to find a place to sit, David stood against the wall alongside his good friend Beth, a fellow second-year. In the room he saw students of all levels, from first-years to fourth-years, engaged in excited chatter.

The third-year class president, Sam, stood up. "Okay everyone, quiet down so that we can begin the discussion. We had not expected a turnout of this magnitude; it's clear that this is an issue many of you feel quite passionately about. The administration has informed us that adopting a pass/fail system will require a majority vote from the student body."

The volume level in the room suddenly increased.

He continued, "So, we hope that this meeting will serve as a lively debate where students on either side of this issue can share their arguments with the voting body."

"Pass/fail is such a great idea," David whispered to Beth.

To his surprise, she disagreed. "I don't think so," Beth replied. "I personally work harder and perform better when I am graded."

One of the third-year petitioners stood up to argue, "Our medical school is known for being one of the most intensely competitive programs in the country. We are already so stressed out—becoming pass/fail would remove an atmosphere of hypercompetition, and that will be a good change for our mental, emotional, and physical well-being." His words were met with applause from some students in the hall.

Another third-year petitioner presented a counterargument. "The majority of our graduating students match with residency programs each year, and most of those match at one of the programs they ranked in their top three. We've done very well with grades—would the same be true if we became pass/fail? Also, those of us

interested in matching into very competitive specialties, such as dermatology, ophthalmology, and surgical specialties are put at a disadvantage since class rank and academic performance are highly regarded by residency directors in these specialties."

David, who himself had a particular interest in going into surgery, looked around the hall and saw a number of students nodding their heads in agreement. Beth nudged him playfully and whispered, "See what I mean?"

Commentary 1
by Bonnie M. Miller, MD

The primary purpose of any grading system is to measure student achievement of established learning objectives. Performance data let individual students know where they stand in the development of needed competencies. Aggregated performance data supply faculty and medical school administration with information about the effectiveness of teaching. A traditional grade stratifies students according to level of achievement and can motivate students, reward effort, and perhaps signify suitability for a potential area of study. A pass/fail grade indicates simply that a student has achieved an expected level of competence, information that is critically important if medical education is to fulfill its obligation to the public.

The ideal grading system would also encourage the development of desirable professional behaviors. Does a traditional grading system encourage students to constantly strive for excellence, a habit that, theoretically, they would maintain when they no longer receive grades? Does a pass/fail system encourage collegiality, collaboration, and teamwork, since no one is disadvantaged by another's success, and mutual benefit can result from sharing. In the case scenario we are commenting on, is Beth correct in fearing a lack of motivation in the absence of grades, or is David justified in his concern about grade-induced hyper-competitiveness?

I believe that concerns about both consequences are justified, but my experience with grading systems suggests that neither is inevitable. Based on our grade-system change at Vanderbilt University earlier in the decade, I believe that elements such as faculty role modeling, selection of teaching strategies, careful and inclusive selection of the qualities that are being assessed, and use of criteria-based grading systems are more important contributors to student evaluation than whether or not letter grades are used.

Faculty Role
Grading systems exist within the larger context of an educational environment that can powerfully mold the professional development of students. If students are hypercompetitive, it is unlikely that the grading system alone creates that behavior. Similarly, if students consistently aim their efforts at minimal passing performance, the environment might lack the ingredients needed to inspire excellence. Regardless of the grading system, medical school faculty and administration should be aware of

the environments they create and monitor them with vigilance to assure that they support the attitudes and behaviors expected of the profession.

In any grading system, faculty members should serve as role models who demonstrate a passion for excellence and a quest for improvement, both in their teaching efforts and their patient-care responsibilities. Role models who strive for excellence, not because of grades but for the good of those they serve, help students move beyond the external rewards that motivated them in their previous endeavors. Whether in teaching teams or in clinical teams, faculty members can also model the collaboration and collegiality that are important for effective, high-quality patient care. Finally, when faculty members care for the well-being and professional growth of their students, they model the compassionate and nurturing attitudes we hope those students will adopt.

Teaching and Course-Management Strategies
Teaching strategies can also ameliorate the potentially negative side effects of a grading system. Many students study best in groups or learn most deeply when they are challenged to teach their peers, and schools with traditional grading systems can actively promote these approaches. Faculty can use course-management systems that allow all students to see the answers to all questions asked, and students can be encouraged to post helpful articles and learning tips. Team-based learning rewards group performance as opposed to individual effort, while creating pressure not to let one's peers down, which discourages the slacking that a pass/fail system might encourage.

Choosing What to Measure
Perhaps the grading system a school uses is less important than the qualities it chooses to grade. Assessment indeed drives learning, and if we feel that the professional development of our students is critical, we should demonstrate that by assessing it. In both science-based and clinical courses, students should be evaluated on their initiative, engagement with and concern for their own learning, interpersonal skills, teamwork skills and collegiality. Schools can devise grading policies, whether pass/fail or traditional, in which failure to demonstrate one of these key attributes can lead to failure in the course, regardless of cognitive achievement.

Criteria-Based Grading
Finally, the use of a normative versus a criteria-based grading system can influence student behaviors. In the former, the grade distribution is determined by comparative student performance, limiting the number of highest grades and creating an atmosphere in which one student's performance can influence the grade of another. This is more likely to induce competition. In a criteria-based system, the requirements for each grade interval are predetermined, and any student who meets the designated requirements receives the designated grade, even if an entire class qualifies for an A. While this model could lead to grade inflation, it does recognize all students who achieve a certain level of excellence. And shouldn't all medical teachers aspire to the goal of having all students excel?

The Vanderbilt Grading Experience

In 2002, Vanderbilt University reexamined its traditional letter grading system. Like students at David and Beth's school, our students performed very well in the residency match, and we were leery of changes that would make it more difficult for program directors to evaluate students. Unlike students at David and Beth's school, ours did not complain of an overly competitive atmosphere. I'd like to think that this was because of our collegial educational environment, but a criteria-based system probably helped. Our greatest concern at that time was for the fairness of grades in the first year of medical school. Because of the wide variation in our students' undergraduate preparation and the difficulties of adjusting to medical school, we felt that letter grades reflected not only effort and ability, but also the strength of the undergraduate program, the major a student had selected, and the ease of social transition. Most of our students who received marginal grades in the first year subsequently performed at very high levels, but were left with transcripts that marred their overall records.

To balance our concern for first-year grades with our concern for the impact of a pure pass/fail system on the residency application process, we decided upon a hybrid system with pass/fail in the first year only; honors/pass/fail in the second year; and honors/high pass/pass/fail in the third and fourth years. We hoped that the noncompetitive culture of collaboration established in the first year would continue throughout the remaining 3 years, even as more grade intervals were introduced.

Some faculty feared, like Beth, that first-year students would lack the motivation to put forth their strongest efforts. Fortunately, this fear never became a significant reality. Our curriculum remains rigorous and demands hard work, and the environment still encourages our students to reach for excellence. Occasionally a student's performance slips on the last exam in a course if he or she is easily within the passing range, but this has not been a large enough effect to diminish overall class performance from year to year. Student performance in the subsequent years of medical school and on Step 1 of the United States Medical Licensing Examination (USMLE) has actually improved, relieving anxieties about the grading system's long-term negative impacts on the learning habits.

Paradoxically, in the first year of the transition, students and faculty sensed an increase in student competitiveness in the second-year class, even though this class entered with a traditionally graded system. We quickly realized that this resulted from a concurrent switch to a normative-based system that limited the number of honors grades to 25 percent of the class. In the following year, we reverted to a criteria-based system that set the honors bar extremely high to combat grade inflation but allowed all students who cleared that bar to receive an honors grade. Many students in that second-year class were also unhappy with the change and reported that they had selected Vanderbilt because of its traditional grading system. We learned from this experience that whenever possible, major policy and curriculum changes should be phased in with the entering classes. I have also become a strong

believer in a criteria-based system that sets high standards but proudly recognizes all students who meet them.

Because we maintained four grading intervals in the clinical years, we experienced no measurable change in the outcomes of our residency match. For schools that use a pass/fail only system throughout the 4-year curriculum, program directors rely more on qualitative measures, such as the comments recorded on clerkships assessment forms, letters of recommendation, and the nature of student leadership and scholarship accomplishments. With a sense that these subjective measures are less reliable than the objectivity of grades, program directors also tend to rely more heavily on Step 1 scores and the reputation of the medical school.

No grading system is perfect in its ability to assess learners accurately, promote professional behaviors, and predict future accomplishments. Regardless of the system selected, a school must be aware of the potential for unintended consequences and should strive for an educational environment that counters these and encourages students to excel for the right reason, which is that their excellence will someday improve the lives of others.

Bonnie M. Miller, MD, is the senior associate dean for health sciences education at Vanderbilt University School of Medicine in Nashville.

Commentary 2
by Adina Kalet, MD, MPH

As medical educators, our responsibility to society is to ensure that all physicians are competent to practice medicine. Ideally, both faculty and students should enthusiastically engage in an evaluation system that facilitates our fulfilling this responsibility. I am a strong believer in a grading system that is ultimately pass/fail—but is at the same time rich in confidential, formative feedback that helps students identify their strengths and weaknesses. To be meaningful, the "pass" thresholds must be competency- and criterion-based, not arbitrary or norm-referenced, i.e., predetermined percentages of students pass and fail.

Competitive residency programs choose residents based on whatever evidence of their abilities exists. Residencies are looking for students who are a good fit for their program, well prepared, and capable of handling the work. The absence of letter grades on the formal transcript, without evidence of a rigorous, reliable assessment process is problematic for two reasons. First, it places enormous, undeserved pressure on students to do well on National Board Exams. Second, this approach overemphasizes the reputation of the medical school and its admissions policies.

The debate presented in the case scenario focuses on the wrong outcomes. For example, students often defend pass/fail systems as more conducive to a relaxed learning environment because there is less interpersonal competition. I am not certain that this reflects reality. All medical students are highly achievement-oriented

and many are competitive by nature. To be successful and competent physicians they must learn to manage the negative impact of these otherwise valuable personal traits in complex and competitive environments. On the other side of the argument, pass/fail systems disadvantage students who are consistently struggling because it allows them to squeak by without being identified for special attention early. In addition, even in schools like mine, NYU Medical Center, that operate with a pass/fail preclinical system, numeric grades are generated and followed for certain purposes (e.g., AOA determination), and students are well aware of this contradictory policy.

In saying that the grades debate often focuses on the wrong outcome, I also mean that scores on exams are only useful if the exams themselves are reliable and valid measures of what they are meant to measure. Ideally, competency exams would provide students with detailed information to help determine whether they had the minimum competency to serve as physicians. We would overcome current weaknesses in measuring the remarkable capacities some students have in areas such as interdisciplinary teamwork and complex critical thinking. Once we have decided on fair, criterion-based measures that assess critical competencies, there is no way we could ethically, morally, or professionally argue against using such measures. Since most of our exams or grading systems do not reach this level of evidence, however, we use them as blunt instruments rather than sources of meaningful information.

In sum, I don't care as much as many students do about whether we use pass/fail or other systems. I care that we measure what is important and act on those measures to ensure excellence in our graduates.

Adina Kalet, MD, MPH, is the Arnold P. Gold Professor of humanism and professionalism and an associate professor of medicine and surgery at New York University School of Medicine. She has a long-standing research interest in assessment of clinical competence and the relationship between medical education and patient outcomes. She has mentored three cohorts of NYU SOM *Virtual Mentor* student editors.

Commentary 3
by Ryan C. VanWoerkom, Nicholas Zorko, and Julia Halsey

During the late 1960s and early 1970s, medical schools moved away from traditional grading systems and began adopting pass/fail or honors/pass/fail evaluation [1]. It is thought that the impetus for these changes originated with the concern that grade-based learning did not prepare for lifelong learning outside of the academic world and that it suppressed creativity and increased stress [1, 2]. On the other hand, it is well-known that residency directors hold the dean's letter in high regard and favor the more discriminative letter-grade evaluation report [1, 3, 4].

The ultimate quick test in medicine is applying the principle of *primum non nocere* (first do no harm). Is there a possibility that by changing the grading system to a less rigorous, more comfortable pass/fail system we may be harming patients? This would occur indirectly by allowing some students to slip through the cracks of a low-demand education and evaluation system. Gonnella et al. noted that students in need of remediation (not meeting basic standards set for competence in medical education) often went unidentified under a pass/fail system. "Failure to identify students who pass only narrowly results in the suppression of information that is critical to the future development of the students, and is important in the prevention of problems in professional practice" [5]. This does not bode well for patients, even if only a few sub-par students slip through the system without undergoing appropriate remediation.

One example of a problem in professional practice could occur while a student or resident is caring for patients on a hospital team. The extra effort spent by one student studying for an "A" may trigger a memory for the correct tests needed to arrive at a diagnosis and implement an alleviating treatment, a connection that another student who only wanted to pass may not have made. The use of pass/fail grading has been correlated by some groups with poorer performance on exams [8, 9]. Additional information supporting this view was found in a study of surgery residents trained under different grading systems in medical school. Moss et al. found that residents who attended medical schools that assigned grades performed better than those who attended schools that used pass/fail systems [6]. Proponents of pass/fail grading argue that students working in such systems report a greater sense of satisfaction and well-being, but there is evidence refuting this reduction in anxiety upon implementation of a pass/fail grading system [7]. This perceived decrease in anxiety, regardless of validity, may not be worth the decrease in knowledge acquisition that may occur with less rigorous study habits.

Students' personal characteristics and attributes may influence their behavior and attitudes as strongly as a strictly graded traditional system with its intense pressure to perform well—the extrinsic factors—but the two are not easily separated. As one comes closer to measuring an extrinsic factor in medical education, he or she inadvertently affects the intrinsic. Consider, for example, the competitiveness that is said to infect medical students. A student who is willing to pull ahead at the risk of alienating classmates may be innately achievement-oriented, so the cause for his or her behavior is independent of the medical school environment and its pressure to compete.

Many schools have opted for the honors/pass/fail grading system, which does not eliminate the pressure or incentive for students who wish to compete for honors grades. Honors/pass/fail may have the paradoxical effect of placing additional pressure on competitive students to perform even better simply because their grading system fails to discriminate adequately.

A survey of surgery clerkship directors revealed consensus that a three-tiered system did not do enough to differentiate students appropriately. Pass/fail programs, this Ravelli et al. study concluded, "produced little reliable discrimination" between the quality of students and their peers [2]. With this in mind, it is more just to acknowledge a continuum of grades properly than to differentiate only between pass/fail. Consider a student who received the all-time top score for a medical school exam and was given the same grade as a student who passed by one question. This system results in general statements of evaluation for a majority of students without providing a means of recognition for outstanding efforts.

Although many medical schools tout their pass/fail grading system as a means of attracting prospective medical students, these same schools, in truth, rank their students because they know that residency programs want them to distinguish among students. If students are not ranked in a traditional numerical order (e.g., 1/125), they are lumped in quartiles. In order for medical schools to maintain clout in placing their students in competitive residencies, the Medical Student Performance Evaluations (MSPEs) that they send to residency programs must rank students in some useful way. This may even lead to confusion among students regarding their own rank systems.

Turning to the other side of the debate—the argument for pass/fail grading—students have more compelling motivators than grades. Having made it through the weeding process in high school and college classes and even the application process where grades were the most important criteria, medical students need to acquire the knowledge necessary to pass the national boards, obtain residencies and fellowship, and establish a satisfying career. At this point in their medical education, they have greater motivators to learn than simply to get an A on a test.

The letter-grading system also suffers from grade-inflation, which has caused distress in admissions committees and employers of various disciplines. Grade inflation has placed a greater significance on standardized testing as the most objective way for schools to compare candidates from different programs. This in turn, may make the medical board exams a more stressful experience.

While much of this discussion may not seem to be directly related to ethics, in the grand scheme of things, performing at a level which is anything less than one's best has the potential to be detrimental to a patient's well-being and is therefore unethical. The AMA *Code of Medical Ethics* states,

> Incompetence, corruption, or dishonest or unethical conduct on the part of members of the medical profession is reprehensible. In addition to posing a real or potential threat to patients, such conduct undermines the public's confidence in the profession [10].

Therefore, medical students' ethical obligation encompasses the duty to prevent incompetence within their profession.

Steve Prefontaine put it best: "To give anything less than your best is to sacrifice the gift." As physicians or future physicians, we owe it to our patients and society to give our absolute best effort in exchange for the trust and responsibility for their lives they have given over to our care. We have been given a gift and privilege to study and practice medicine and should thus handle it appropriately regardless of the method used to evaluate us.

References
1. Dietrick JA, Weaver MT, Merrick HW. Pass/fail grading: a disadvantage for students applying for residency. *Am J Surg.* 1991;162:(1)63-66.
2. Ravelli C, Wolfson P. What is the "ideal" grading system for the junior surgery clerkship? *Am J Surg*. 1999;177(2):140-144.
3. Lurie SJ, Lambert DR, Grady-Weliky TA. Relationship between dean's letter rankings and later evaluations by residency program directors. *Teach Learn Med*. 2007:19(3):251-256.
4. Provan JL, Cuttress L. Preferences of program directors for evaluation of candidates for postgraduate training. *CMAJ*. 1995;153(7):919-923.
5. Gonnella JS, Erdmann JB, Hojat M. An empirical study of the predictive validity of number grades in medical school using 3 decades of longitudinal data: implications for a grading system. *Med Educ*. 2004;38(4):425-434.
6. Moss TJ, Deland EC, Maloney JV Jr. Selection of medical students for graduate training: pass/fail versus grades. *N Engl J Med*. 1978;299(1):25-27.
7. Yarbro RC. A comparison of anxiety levels of students taking pass/fail versus grade in student teaching. *Tenn Educ*. 1982;12(2):33-36.
8. Weller LD. The grading nemesis: an historical overview and a current look at pass/fail grading. *J Res Devel Educ*. 1983;17:39-45.
9. Suddick DE, Kelly RE. Effects of transition from pass/no credit to traditional letter grade system. *J Exp Educ*. 1981;50:88-90.
10. American Medical Association. Opinion 9.04. Discipline and medicine. Code of Medical Ethics. Chicago, IL: American Medical Association. 1994. http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion904.shtml. Accessed October 12, 2009.

Ryan C. VanWoerkom, is a fourth-year medical student at the University of Utah in Salt Lake City, with plans to enter a career in internal medicine. He serves as the chair of the Committee on Bioethics and Humanities for the American Medical Association-Medical Student Section as well as being the Midwest representative to the American College of Physicians Council of Student Members.

Nicholas Zorko is a fourth-year MD/PhD student at The Ohio State University in Columbus. He graduated from Ohio State with a bachelor's degree in biology in 2006, and is currently the vice chair for the Committee on Bioethics and Humanities for the American Medical Association-Medical Student Section.

Julia Halsey is a third-year medical student at the University of Missouri in Columbia. She graduated from Truman State University in Kirksville, Missouri, with

a bachelor's degree in biology and from Trinity International University in Deerfield, Illinois, with a master's degree in bioethics. She currently serves as the student representative to the AMA's Council on Ethical and Judicial Affairs.

Related in VM
Nurturing Leaders for an Environment of Change, November 2009

Is There More to Medical School than Grades? December 2003

The people and events in this case are fictional. Resemblance to real events or to names of people, living or dead, is entirely coincidental.

The viewpoints expressed on this site are those of the authors and do not necessarily reflect the views and policies of the AMA.