

AAIM Perspectives

AAIM is the largest academically focused specialty organization representing departments of internal medicine at medical schools and teaching hospitals in the United States and Canada. As a consortium of five organizations, AAIM represents department chairs and chiefs; clerkship, residency, and fellowship program directors; division chiefs; and academic and business administrators as well as other faculty and staff in departments of internal medicine and their divisions.

The Grade Debate: Evidence, Knowledge Gaps, and Perspectives on Clerkship Assessment Across the UME to GME Continuum



Katherine R. Schafer, MD,^a Lonika Sood, MBBS, MHPE,^b Christopher J. King, MD,^c
Irene Alexandraki, MD, MPH,^d Paul Aronowitz, MD,^e Margot Cohen, MD, MEd,^f Katherine Chretien, MD,^g
Amit Pahwa, MD, SFHM,^g E Shen, PhD,^a Donna Williams, MD,^a Karen E. Hauer, MD, PhD^h

^aDepartment of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC; ^bElson S. Floyd College of Medicine, Washington State University, Spokane; ^cDivision of Hospital Medicine, Department of Medicine, University of Colorado School of Medicine, Aurora; ^dUniversity of Arizona College of Medicine-Phoenix; ^eUniversity of California Davis School of Medicine; ^fDepartment of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia; ^gJohns Hopkins University School of Medicine, Baltimore, Md; ^hUniversity of California, San Francisco.

INTRODUCTION

Core clerkships are typically a medical student's first fully immersive clinical experiences where they learn to work in teams and contribute to patient care across specialties. Assessment of student clinical performance in the core clerkships can serve many purposes, including providing feedback that guides learning, ensuring achievement of competencies defined by each medical school, and determining readiness for advancement in the curriculum. With the increasingly competitive nature of the residency match, the purpose of assessment for determining grades may supersede the initially intended purposes for both students and educators. Although most core clerkships use multi-tiered rather than binary pass/fail grading systems

(henceforth referred to as “tiered” and “pass/fail” for simplicity), prior studies have raised concerns about how bias, lack of fairness, imprecision, and variability in clerkship grading across institutions can negatively affect students and threaten the accuracy of this information to inform residency selection.¹⁻⁴ In response to these concerns, some medical schools in the United States have transitioned to pass/fail grading in the core clerkships. Data from the 2018 Clerkship Directors in Internal Medicine survey indicated that most medical schools use tiered grades, and only 4.6% had adopted a pass/fail grading system in their internal medicine clerkship.⁵ Association of American Medical Colleges (AAMC) data from 2021 indicated that 16% of medical schools had adopted a pass/fail system across all required clinical clerkships, suggesting a potential trend toward pass/fail clerkship grading, with the caveat that the COVID-19 pandemic may have spurred temporary changes.²

Clarity on the best practices for clerkship assessment and grading is particularly important within the context of the change in the reporting of the US Medical Licensing Examination (USMLE) Step 1 to pass/fail, continued residency application inflation, and

Funding: None.

Conflict of Interest: None.

Authorship: All authors had access to the data and a role in writing this manuscript.

Requests for reprints should be addressed to Katherine R. Schafer, Department of Internal Medicine, Inpatient Internal Medicine Clerkship Director, Wake Forest University School of Medicine, 1 Medical Center Boulevard, Winston-Salem, NC 27157.

E-mail address: kschafer@wakehealth.edu

virtual interviewing for residency positions. Data are lacking with regard to views on grading across the continuum from undergraduate medical education (UME) to graduate medical education (GME) to inform a unified approach. This paper describes the current state of knowledge about clerkship grading and examines perspectives from medical educators within internal medicine, spanning the UME to GME spectrum. We aim to generate dialogue and identify workable approaches to address this ongoing challenge in assessment and grading.

THE CASE FOR PASS/FAIL GRADING

Wellness

In pre-clerkship curricula, pass/fail grading promotes student well-being without negatively affecting learning outcomes as measured by licensing examination performance.^{6,7} Capitalizing on this evidence, the majority of US medical schools use pass/fail grading in pre-clerkship courses, but most use tiered grading systems in the core clerkships.² However, in core clerkships, pass/fail grading has been shown to improve student perceptions of wellness and agency to make decisions about their learning based on their own reflection and goal-setting, such as when to study or see a patient.⁸ Students perceive the clinical learning environment with pass/fail grading to be mastery-oriented rather than performance-oriented and less competitive than with tiered grading systems.⁹

Learner Motivation

A mastery orientation aligns with learners' intrinsic motivations to seek feedback from supervisors, take on new learning opportunities, and view their shortcomings or gaps as additional learning opportunities. It is associated with a growth mindset, which is beneficial for a career that requires learning and adapting to new information and practices.¹⁰⁻¹² In contrast, grades and comparisons to peers promote a performance orientation, in which students are motivated by external rewards (grades) and validation.⁴ The pressure to earn grades to be competitive for residency arises for students during a particularly foundational period in their medical training. Consequently, students are prompted to seek praise and reinforcement, fearing critical feedback as a threat to their grades despite its usefulness to drive learning. They view supervisors as arbiters of grades rather than resources for learning. In fact, students perceive that

"being liked" is most important while "rapport with patients and families" and "improvement" are least important in determining tiered grades.⁴

Equity

Clerkship grading is vulnerable to bias. Studies from multiple institutions have demonstrated group differences in clerkship assessments and grades favoring students from backgrounds not underrepresented in medicine.^{1,13,14} Even small differences in ratings from supervisors of students can produce large differences in grades and selection to the Alpha Omega Alpha Honor Society (AOA).¹³ A study of over 50,000 US medical students utilizing AAMC data found that students who were lesbian, gay or bisexual, low-income, or non-white were underrepresented in AOA membership.¹⁵

Inequity in assessment can arise from multiple sources, including different opportunities to see patients and present to the team, microaggressions targeting students, imposter syndrome and stereotype threat, and biased assessments by individual supervisors.¹⁶ To promote equity, educators designing assessment systems and practices must first adhere to principles of good practice in assessment.¹⁷ Frequently observing learners and providing feedback in a safe, low-stakes manner allows all students to practice and improve prior to high-stakes assessments.¹⁸ Training faculty about implicit bias, monitoring data for differential outcomes, and acting upon those results, also mitigates inequity.

Grade Variability

There is marked variation and imprecision of clerkship grading across medical schools in the United States. The term "honors" means very different things from school to school with the percentage of students receiving honors in various clerkships ranging from 2% to 93% —a remarkable discrepancy indicating that there is little, if any, standardization of what this designation means within and across schools.¹⁹ Determining clerkship grades can be challenging due to variability in assessment among student supervisors and settings and lack of a frame-of-reference for performance expectations.^{19,20} Variability also exists among medical schools in how they weigh National Board of Medical Examiners subject exams, student clinical performance, and personal attributes in grade assignment.^{5,19,21} Despite these variable and imprecise grading labels and methods of determining grades,

PERSPECTIVES VIEWPOINTS

- Student assessment in core clerkships guides learning and the achievement of expectations. However, tensions persist on the role of grading.
- Strategies to balance the goals of feedback and learning, equity, learner motivation, and wellness with the ability to discriminate between student performance remain elusive.
- It is imperative that stakeholders across the UME-GME continuum partner to determine the best practices in assessment and grading.

grades are heavily relied upon as surrogate markers of student quality when it comes to residency applicant screening and ranking.

THE CASE FOR TIERED GRADING

Student Performance Discrimination

Fundamentally, an ideal grading system should measure meaningful differences among students at different levels of academic attainment.²² Durning and Hemmer indicated that the purpose of grading is to summarize the student's overall performance, concisely distinguish various levels of performance, and identify students who are likely to succeed in progressively more challenging educational environments.²³ A pass/fail grading system does not differentiate levels of student performance, which can negatively impact top performing students and miss identifying students who need remediation, particularly if there is not other information such as narrative evaluations to distinguish among students.²² Additionally, shifting the focus of assessment away from grade assignments requires significant faculty development and culture change to support the increase in formative feedback and use of narrative or other assessment data recommended within pass/fail systems.^{3,8}

Despite concern that tiered grading systems do not correlate with future performance, one study showed a strong association between core clerkship grades and internship ratings by program directors in professionalism and knowledge.²⁴ Data from a 7-year cohort of graduates from the Uniformed Service University demonstrated that internal medicine clerkship grades correlate with American Board of Medical Specialties certification. Students with higher internal medicine clinical points (eg, summation of points received from faculty during the internal medicine clerkship) and internal medicine total points (eg, summation of clinical points and examination points from National Board of Medical Examiners and in-house exams) were more likely to achieve successful board certification.²⁵

Communication of Evaluation of Student Achievement to Residency Programs

Residency program directors rely on clerkship grades as one of the key factors used to select applicants for postgraduate positions.²⁶ Over time, the application process has become more competitive as applicants apply to increasing numbers of programs each year, particularly with the advent of virtual interviews. In 2021, the average US medical graduate (MD and DO) applying to categorical internal medicine applied to 46 programs, while international medical graduates applied to 98.²⁷ On the receiving end, categorical internal medicine residency programs received an average of 3135 applications to review in 2021, which was up

from 2708 in 2020.²⁷ Given the limited resources, programs must identify efficient and equitable processes to review an increasing number of applications.

A survey of internal medicine program directors in 2017 suggested that USMLE Step 2 clinical knowledge (CK), internal medicine clerkship grade, and USMLE Step 1 were the most commonly used criteria for offering interviews.²⁸ In 2021, internal medicine program directors noted that USMLE Step 2 CK and clerkship grades would be important factors for selection of applicants for interviews with the transition of USMLE Step 1 to pass/fail reporting.²⁹ Students who do not have transcripts with tiered clerkship grades may be subjected to even higher pressure to perform well on the high stakes USMLE Step 2 CK exam. Multiple studies have attempted to uncover which residency selection criteria are predictive of residency performance in internal medicine; clerkship grades consistently correlate with measures of physician performance, including ACGME milestones and licensure examinations.^{30,31}

Equity, Stress, and Well-Being

Although some studies have shown that a pass/fail system may have a positive impact on student well-being, longitudinal data and evidence that these benefits occur during the period of learning are lacking.^{6,32,33} Furthermore, it is important to attend to whether tiered grades should be eliminated completely or supplanted by other, similar tiered systems. Medical schools with pass/fail systems may use other quantitative measures to differentiate students. In a study by Bloodgood et al, although a pass/fail system existed in the medical school's pre-clerkship curriculum, cumulative honors were awarded to 20% of the class at the end of the second year based on average of percentage scores in all courses.³³ Using other quantitative metrics in a pass/fail system can shift student stress from one environment to another and potentially jeopardize any benefits on student well-being.³⁴

Overall, available data on pass/fail grading systems in the core clerkships are limited. A qualitative study of student perceptions around pass/fail grading at one medical school revealed students were supportive of this change for multiple reasons including equity and well-being, although some students did question their ability to stand out among their colleagues without grades.⁷ The ability to stand out may be especially important to students who are applying to highly competitive specialties and may displace stress to other areas, such as USMLE Step 2 CK and the need to pursue away rotations during the final year of medical school.

PERSPECTIVES

Tensions regarding how core clerkship assessment can support student growth and skill development while

also supporting the residency selection process remain across the UME-GME continuum.^{35,36} On the one hand, those who support pass/fail grading suggest that tiered grades may shift student priorities away from patient-centered and personal improvement-focused learning toward appearing capable and knowledgeable to their supervisors. Proponents of tiered grading counter that a purely binary grading system hinders discriminating student performance thus limiting the ability to identify under-performing learners and impacting recruitment to residency programs.

Data exist for and against a pass/fail grading system, but significant gaps remain in the evidence base to support one approach definitively over another. Some educators suggest focusing more on narrative comments rather than grades as supervisors' narrative comments do provide information that educators can use to distinguish between student performances.³⁷ Unfortunately, even narrative descriptions of clerkship performance can demonstrate bias. In a study of words used to describe student performance in clerkships, though most words were used similarly to describe students with different genders or race/ethnicities, there were subtle word choice differences that disfavored women and students who self-identified as an under-represented minority through faint praise and more focus on effort and collaboration than skill and achievement.³⁸ The debate over grading extends beyond medical education and may offer suggestions that medical educators across the internal medicine continuum may implement.

Existing research has analyzed the pros and cons of clerkship grading from the perspective of particular stakeholder groups including students and clerkship and program directors.^{3,4,8,39} These findings can position one stakeholder group in opposition to another with advocacy for the needs of students against the needs of residency selection committees. Data supporting pass/fail grades are largely qualitative and limited to schools whose "name recognition" may not be generalizable across the country. There are currently no randomized, multicenter, or prospective studies assessing the impact of different grading systems on longitudinal outcomes. These types of investigation would be complex and resource-intensive.

NEXT STEPS

The ultimate goal of medical education is to produce competent physicians who can provide quality, unsupervised care for patients. In order to determine best practices for learner assessments, the UME to GME continuum will need to collaborate to undertake robust, multi-institution studies such as iCOMPARE (randomized controlled trial investigating flexible versus standard duty hours in residency).⁴⁰ The next steps should

build upon the Coalition for Physician Accountability UME-GME Review Committee recommendations by:

- 1) Rigorously investigating the ability of different grading systems to measure educational outcomes, prioritize equity, and impact student well-being. This step could begin with retrospective data from the AAMC and a diverse, representative sample of institutions.
- 2) Developing pilot studies of new assessment methods; correlates of interest could include fourth-year performance, Step 2 CK exam scores, AAMC Graduation Questionnaire responses, and Internal Medicine Board certification.
- 3) UME and GME stakeholders utilizing these data to develop principles for improving assessment and grading procedures to optimize the benefit for all stakeholders and then reach consensus on assessment best practices.⁴¹

Throughout this work, attention to equity is critical so that assessment practices afford all learners the opportunities to learn, grow, and achieve competence for practice. Some key groundwork already exists on identifying prioritized educational outcomes of interest and ways to optimize equity in assessment across the learner transitions.^{42,43}

References

1. Low D, Pollack SW, Liao ZC, et al. Racial/ethnic disparities in clinical grading in medical school. *Teach Learn Med* 2019;31(5):487–96.
2. Association of American Medical Colleges. Grading systems used in medical school programs. Available at: <https://www.aamc.org/data-reports/curriculum-reports/interactive-data/grading-systems-used-medical-school-programs>. Accessed October 20, 2022, 2022.
3. McDonald JA, Lai CJ, Lin MYC, O'Sullivan PS, Hauer KE. "There is a lot of change afoot": a qualitative study of faculty adaptation to elimination of tiered grades with increased emphasis on feedback in core clerkships. *Acad Med* 2021;96(2):263–70.
4. Bullock JL, Lai CJ, Lockspeiser T, et al. In pursuit of honors: a multi-institutional study of students' perceptions of clerkship evaluation and grading. *Acad Med* 2019;94(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 58th Annual Research in Medical Education Sessions):S48–56.
5. Hernandez CA, Daroowalla F, LaRochelle JS, et al. Determining grades in the internal medicine clerkship: results of a national survey of clerkship directors. *Acad Med* 2021;96(2):249–55.
6. Spring L, Robillard D, Gehlbach L, Simas TA. Impact of pass/fail grading on medical students' well-being and academic outcomes. *Med Educ* 2011;45(9):867–77.
7. Reed DA, Shanafelt TD, Satele DW, et al. Relationship of pass/fail grading and curriculum structure with well-being among pre-clinical medical students: a multi-institutional study. *Acad Med* 2011;86(11):1367–73.
8. Seligman L, Abdullahi A, Teherani A, Hauer KE. From grading to assessment for learning: a qualitative study of student perceptions surrounding elimination of core clerkship grades and enhanced formative feedback. *Teach Learn Med* 2021;33(3):314–25.
9. Bullock JL, Seligman L, Lai CJ, O'Sullivan PS, Hauer KE. Moving toward mastery: changes in student perceptions of clerkship

- assessment with pass/fail grading and enhanced feedback. *Teach Learn Med* 2022;34(2):198–208.
10. Dweck C. Harvard Business Review. What having a “growth mindset” actually means. Available at: <https://hbr.org/2016/01/what-having-a-growth-mindset-actually-means#:~:text=Everyone%20is%20actually%20a%20mixture%20of%20fixed%20and,mindset%20is%20just%20about%20praising%20and%20rewarding%20effort>. Accessed November 18, 2022.
 11. Dhaliwal G, Hauer KE. Excellence in medical training: developing talent-not sorting it. *Perspect Med Educ* 2021;10(6):356–61.
 12. Richardson D, Kinnear B, Hauer KE, et al. Growth mindset in competency-based medical education. *Med Teach* 2021;43(7):751–7.
 13. Teherani A, Hauer KE, Fernandez A, King TE, Lucey C. How small differences in assessed clinical performance amplify to large differences in grades and awards: a cascade with serious consequences for students underrepresented in medicine. *Acad Med* 2018;93(9):1286–92.
 14. Colson ER, Pérez M, Blaylock L, et al. Washington University School of Medicine in St. Louis case study: a process for understanding and addressing bias in clerkship grading. *Acad Med* 2020;95(12S Addressing harmful bias and eliminating discrimination in health professions learning environments):S131–5.
 15. Hill KA, Desai MM, Chaudhry SI, et al. Association of marginalized identities with Alpha Omega Alpha Honor Society and Gold Humanism Honor Society membership among medical students. *JAMA Netw Open* 2022;5(9):e2229062.
 16. Lucey CR, Hauer KE, Boatright D, Fernandez A. Medical education’s wicked problem: achieving equity in assessment for medical learners. *Acad Med* 2020;95(12S Addressing harmful bias and eliminating discrimination in health professions learning environments):S98–S108.
 17. Norcini J, Anderson MB, Bollela V, et al. 2018 consensus framework for good assessment. *Med Teach* 2018;40(11):1102–9.
 18. Fyfe M, Horsburgh J, Blitz J, Chiavaroli N, Kumar S, Cleland J. The do’s, don’ts and don’t knows of redressing differential attainment related to race/ethnicity in medical schools. *Perspect Med Educ* 2022;11(1):1–14.
 19. Alexander EK, Osman NY, Walling JL, Mitchell VG. Variation and imprecision of clerkship grading in U.S. medical schools. *Acad Med* 2012;87(8):1070–6.
 20. Zaidi NLB, Kreiter CD, Castaneda PR, et al. Generalizability of competency assessment scores across and within clerkships: how students, assessors, and clerkships matter. *Acad Med* 2018;93(8):1212–7.
 21. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;357(9260):945–9.
 22. 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):425–34.
 23. Durning SJ, Hemmer PA. Commentary: grading: what is it good for? *Acad Med* 2012;87(8):1002–4.
 24. Greenburg DL, Durning SJ, Cohen DL, Cruess D, Jackson JL. Identifying medical students likely to exhibit poor professionalism and knowledge during internship. *J Gen Intern Med* 2007;22(12):1711–7.
 25. Durning SJ, Dong T, Hemmer PA, et al. Are commonly used premedical school or medical school measures associated with board certification? *Mil Med* 2015;180(4 Suppl):18–23.
 26. Cullen MW, Reed DA, Halvorsen AJ, et al. Selection criteria for internal medicine residency applicants and professionalism ratings during internship. *Mayo Clin Proc* 2011;86(3):197–202.
 27. Association of American Medical Colleges. ERAS statistics. Available at: <https://www.aamc.org/data-reports/interactive-data/eras-statistics-data>. Accessed March 30, 2022.
 28. Angus SV, Williams CM, Stewart EA, Sweet M, Kisielewski M, Willett LL. Internal medicine residency program directors’ screening practices and perceptions about recruitment challenges. *Acad Med* 2020;95(4):582–9.
 29. Mun F, Scott AR, Cui D, Chisty A, Hennrikus WL, Hennrikus EF. Internal medicine residency program director perceptions of USMLE Step 1 pass/fail scoring: a cross-sectional survey. *Medicine (Baltimore)* 2021;100(15):e25284.
 30. Kenny S, McInnes M, Singh V. Associations between residency selection strategies and doctor performance: a meta-analysis. *Med Educ* 2013;47(8):790–800.
 31. Golden BP, Henschen BL, Liss DT, Kiely SL, Didwania AK. Association between internal medicine residency applicant characteristics and performance on ACGME milestones during intern year. *J Grad Med Educ* 2021;13(2):213–22.
 32. Rohe DE, Barrier PA, Clark MM, Cook DA, Vickers KS, Decker PA. The benefits of pass-fail grading on stress, mood, and group cohesion in medical students. *Mayo Clin Proc* 2006;81(11):1443–8.
 33. Bloodgood RA, Short JG, Jackson JM, Martindale JR. A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Acad Med* 2009;84(5):655–62.
 34. Guo R, Muir J. Toronto’s honours/pass/fail debate: the minority position. *Univ Toronto Med J* 2008;85(2):95–8.
 35. Govaerts MJB, van der Vleuten CPM, Holmboe ES. Managing tensions in assessment: moving beyond either-or thinking. *Med Educ* 2019;53(1):64–75.
 36. Hauer KE, Williams PM, Byerley JS, Swails JL, Barone MA. Blue skies with clouds: envisioning the future ideal state and identifying ongoing tensions in the UME-GME transition. *Acad Med* 2022.
 37. Hatala R, Sawatsky AP, Dudek N, Ginsburg S, Cook DA. Using in-training evaluation report (ITER) qualitative comments to assess medical students and residents: a systematic review. *Acad Med* 2017;92(6):868–79.
 38. Rojek AE, Khanna R, Yim JWL, et al. Differences in narrative language in evaluations of medical students by gender and under-represented minority status. *J Gen Intern Med* 2019;34(5):684–91.
 39. Wang A, Karunungan KL, Shlobin NA, et al. Residency program director perceptions of resident performance between graduates of medical schools with pass/fail versus tiered grading system for clinical clerkships: a meta-analysis. *Acad Med* 2021;96(11S):S216–7.
 40. Desai SV, Asch DA, Bellini LM, et al. Education outcomes in a duty-hour flexibility trial in internal medicine. *N Engl J Med* 2018;378(16):1494–508.
 41. The Coalition for Physician Accountability’s Undergraduate Medical Education-Graduate Medical Education Review Committee. Recommendations for Comprehensive Improvement of the UME-GME Transition. Available at: <https://physicianaccountability.org/wp-content/uploads/2021/08/UGRC-Coalition-Report-FINAL.pdf>. Accessed November 18, 2022.
 42. Pincavage AT, Osman NY, Alexandraki I, et al. AAIM recommendations to promote equity in the clerkship clinical learning environment. *Am J Med* 2022;135(8):1021–8.
 43. Lewis K, O V, Garber AM, et al. AAIM recommendations to improve learner transitions. *Am J Med* 2022;135(4):536–42.