Core Clerkship Grading: The Illusion of Objectivity

Karen E. Hauer, MD, PhD, and Catherine R. Lucey, MD

Abstract

Core clerkship grading creates multiple challenges that produce high stress for medical students, interfere with learning, and create inequitable learning environments. Students and faculty alike succumb to the illusion of objectivity—that quantitative ratings converted to grades convey accurate measures of the complexity of clinical performance.

Clerkship grading is the first high-stakes assessment within medical school and occurs just as students are newly immersed full-time in an environment in which patient care supersedes their needs as learners. Students earning high marks situate themselves to earn entry into competitive residency programs and selective specialties. However, there is no commonly accepted standard for how to assign clerkship grades, and the process is vulnerable to imprecision and bias. Rewarding learners for the speed with which they adapt inherently favors students who bring advantages acquired before medical school and discounts the goal of all learners achieving competence.

The authors propose that, rather than focusing on assigning core clerkship grades, assessment of student performance should incorporate expert judgment of learning progress. Competency-based medical education is predicated on the articulation of stepwise expectations for learners, with the support and time allocated for each learner to meet those expectations. Concurrently, students should ideally review their own performance data with coaches to self-assess areas of relative strength and areas for further growth. Eliminating grades in favor of competency-based assessment for learning holds promise to engage learners in developing essential patient care and teamwork skills and to foster their development of lifelong learning habits.

Downloaded from https://journals.lww.com/academicmedicine by BhDMf5ePHKav1zEoum1tQfN4a+kJLhEZgbsIHo4XMi0hCywCX1AWnYQp/IlQrHD3i3D0OdRyi7TvSFl4Cf3VC4/OAVpDDa8K2+Ya6H515kE= on 08/07/2020

Grades are a universally accepted and seemingly objective measure of accomplishment from kindergarten through college. High grade point averages are frequently instrumental to a student's ability to earn entrance into college, graduate, or professional school. It is therefore not surprising that, in the medical education context, residency programs have developed a reliance on core clerkship grades to select the graduating medical students they would most like to enter their programs. The rationale behind the use of core clerkship grades is that they, like the average of test scores in a calculus course, are a measure of a student's achievement and the likelihood of future similar performance by that student in similar circumstances. Moreover, there is an assumption that it is fair, equitable, and necessary to use grades to rank one student in comparison with another. Consequently, clerkship learning has transformed into a high-stakes endeavor because the grade a student earns may determine whether that student can pursue a given specialty or enter graduate training at a certain institution. This reliance on core clerkship grades remains true today, despite the national trend to begin core clerkships in year two of medical school, meaning that a student's aptitude for future success in a given specialty may be judged before the halfway point in his or her medical school education.

Please see the end of this article for information about the authors.

Correspondence should be addressed to Karen E. Hauer, UCSF, 533 Parnassus Ave., U80, Box 0710, San Francisco, CA 94143; telephone: (415) 502-5475; e-mail: karen.hauer@ucsf.edu.

Acad Med. 2019;94:469–472.
First published online August 14, 2018
doi: 10.1097/ACM.0000000000002413
Copyright © 2018 by the Association of American Medical Colleges

How Core Clerkship Grades Miss the Mark

Complexity

Although grades are useful in many educational contexts, core clerkships are not calculus classes. The complexity of competencies a student is expected to master and the variability of contexts (e.g., differences in patient care assignments, peer support, workload, teaching ability of residents and faculty, systems issues) in which they are expected to learn do not translate readily to grades. Designations of "honors" or "pass" do not come close to capturing the nuance of students' performance. And yet, students and faculty alike succumb to the illusion of objectivity—that quantitative ratings converted to grades convey accurate measures of students' performance in complex clinical environments. Although medical schools across the country commonly use pass/fail grading early in the curriculum to encourage student well-being and enhance learning, the same schools assign summative clerkship grades of "honors," "high pass," and "pass" to differentiate students.[1] The availability of grades encourages residency programs routinely to use either number of honors in core clerkships or class rank to compare students within and across medical schools, though there is no standard approach to assigning grades or establishing class rank in U.S. medical schools.[2,3]

Flawed data

In addition to the illusion of objectivity that is communicated by distillation of complex performance into a single grade, there is ample evidence that the data collected to support that grade are frequently flawed. Supervising faculty and residents routinely rate students' skills including history taking, physical examination, and patient communication, although students report rarely being observed performing these

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

clinical skills with patients.[4] Presumably, supervisors rate these skills by inferring from proxy information, such as oral presentation skills during rounds, or seek input from others who have observed the student. Students engage in written work to demonstrate their data collection and reasoning skills, yet rarely receive feedback on their notes or assessments and plans.[5]

The rotational model of medical training exacerbates the problem of assessment as students supervised by faculty who rotate frequently are increasingly challenged to find opportunities to receive observation-based feedback and to demonstrate their ability to respond to feedback.[3,6] Even when faculty do observe students, the ways that they translate those observations to ratings and understand performance expectations are widely variable. Literature on rater cognition urges caution in oversimplifying the complex processes by which faculty make ratings, which are inherently informed by their individual perspectives and biases as well as the specific aspects of particular student–patient encounters.[7]

Narrative evaluations of student performance suffer shortcomings as well. Although narrative feedback on performance is essential for learning and improvement, students report rarely receiving useful narrative feedback during clerkships. In summative evaluations, which do typically contain narrative information, mention of areas for improvement are considered to signal bad performance. Narratives can be replete with "code" language that supervisors learn on the job but that students find obscure and ambiguous.[8]

Unequal opportunities

Grading pressures in core clerkships are not felt equally by all students. In a time-challenged, normatively graded environment, students who adapt quickly and perform well from the start may be rated more favorably than other students, even if they all eventually achieve expected competency. This reality inherently favors students who bring advantages acquired before medical school. Students who come from families of physicians or other health providers, and those who have medical or other professional experience, already understand some of the language and norms of the clinical environment. In reality, they are not "quick learners" but, rather, learners who already know how to fit in.

Students whose experience or cultural background differs from the professional environment of clinical practice may naturally need more time to acclimate. Although these students who have more to learn may apply great effort and learn at a rapid rate, their assessors frequently rate them as lower-performing than some of their peers. Students can learn at different rates and get to the same point, and the ability to demonstrate learning and improvement may be more important in practice. In an environment where extroversion and confidence may be commended as indicators of a stronger performance, students who are naturally quiet or who are more introverted in their thought processes may be viewed as less competent, regardless of their other skills.

Particularly concerning are the ways that current assessment and grading practices in core clerkships may disadvantage students who are underrepresented in medicine. Population group differences in standardized exam performance are known to exist between minority and majority populations and are attributed to long-standing structural racism. Grading paradigms that give substantial weight to standardized exams over clinical performance may amplify the impact of these structural issues. Together, these circumstances contribute to less opportunity for students from groups underrepresented in medicine to earn honors grades; be selected for Alpha Omega Alpha; and compete successfully for highly selective specialties, residency programs, and faculty careers.[2,9]

Shifting the Focus

The primary responsibility of medical school is not to rank students but to ensure that every graduate is prepared with the requisite competencies needed to transition into the next phase of education, where they will be expected to provide patient care and engage in learning with progressively less direct supervision. It is here that our need to assign grades to provide residency programs with information that carries the illusion of objectivity may be at odds with our commitment to ensure that every student achieves the highest possible level of competency. When students feel compelled to earn an honors grade to maintain their career options, they focus on performing rather than on learning and improving, they dread rather than welcome questions that stretch their minds, they fear rather than seek out corrective feedback, they compete rather than collaborate with peers, and they feel stressed rather than energized by the opportunity to learn from their patients.[10] When clerkships attempt to increase the objectivity of grade assignments by relying on standardized (shelf) exams, students prioritize studying for the test over optimizing their direct patient care and interprofessional team skills. When time-challenged faculty view their assessment role as determining whether or not the student is an honors candidate, they focus more on judging immediate performance rather than coaching for future success. All of these maladaptive behaviors are understandable, and all interfere with the culture of collaboration, curiosity, continuous learning, and competency development to which we aspire.

A new approach to assessment

It is time for medical schools to refocus their efforts on designing and implementing programs of assessment that prioritize measuring and continuously advancing student competency development over the assignment of grades that are, at best, imprecise measures of complex performance (Table 1). Emerging paradigms of competency-based medical education and programmatic assessment for, rather than of, learning can guide efforts to communicate important information about student learning and readiness for graduate medical education.[11]

Assessment for learning entails frequent observation and feedback that promote students' development, reflection, and lifelong learning skills. Students should be rewarded for demonstrating learning and improvement, rather than only for performing well when someone is watching.[12] Educators must work with students and students' other educators from the beginning of a core clerkship to incorporate understanding of each

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

Table 1
Recommendations for Core Clerkship Assessment and Grading

Aspect of assessment	Current procedures	Future vision
Purpose of clerkship evaluation and grading	To classify students	To promote learning and development
Clarity of expectations	Opaque, confusing	Transparent, understandable
Feedback	Often misaligned with summary evaluations; high stakes	Frequent, immediate, actionable
Learning progress	Time based	Individualized, based on milestones
Nature of learning context	Frequent changes in team, service, specialty	Continuity with peers, supervisors, setting, patients, and/or team
Assessment tools	Few tools, used mostly for summative assessment, assessment of knowledge as a priority	Multiple tools, frequent formative assessment; assessment of all competency domains a priority
Data that inform performance decisions	Inference based on oral presentations, limited direct observation of patient care	Frequent direct observation of students with patients
Student role in assessment	Passive	Active partner

student's starting point into the learning and assessment process. "Feed forward," in which performance information about a student from one clerkship is provided to the next clerkship, should be embraced as a strategy for longitudinal learning rather than feared by students as information that biases their teachers. Educational efforts across the continuum can build longitudinal performance tracking into learners' experience.[13] Opportunities for students to demonstrate growth and development will be enhanced by structuring at least some longitudinal experiences with supervisors and sites during the core clerkships.[14,15] Fostering the habit of mind of assessment for continuous learning can better prepare students for their obligations of lifelong learning and competency demonstration that underpin our contract with society.

Rather than focusing on assigning grades in the core clerkships, student performance review should incorporate expert judgment of learning progress. For example, a committee of clerkship educators can review assessment data for students, including both quantitative scores and narrative comments, within and across clerkships. Students could simultaneously review their own performance data to self-assess areas of relative strength and areas for further growth. This approach would enable students to focus on learning rather than performance during the core clerkships, so that they are ready to demonstrate their best performance during final-year clinical rotations, including subinternships. There is an argument to be made for returning to grading during final-year rotations because those rotations reflect the most current assessment of each student's ability to perform well in residency.

The effect on residency selection

This proposal will generate questions regarding residency selection. Solving the problem of the transition to residency, which is currently fraught with overwhelming numbers of applications burdening both students and residency programs and a lack of trust between educators on both ends of the educational continuum, will require more than just a change to clerkship grading. However, strategies that begin to address this vexing transition problem are sorely needed, through partnership among all stakeholders involved in undergraduate and graduate medical education and licensure.

Medical schools and residency program directors share the same goal: to match students to the residency programs in which they will provide outstanding patient care while advancing their knowledge and skills. The best information to achieve this goal is found in the combination of grades, narratives, and letters of recommendation from subinternships that describe students' most recent competency achievements. Core clerkship narratives that portray honest, forthright descriptions of how students approached their learning, worked with team members, and incorporated feedback to improve their skills would provide valuable information about multiple skills relevant to residency directors, including professionalism, communication, help seeking, and responsiveness to feedback. Presenting this information visually, such as in a single-page dashboard, could help residency programs efficiently identify students likely to succeed in their program. Educators should advocate for national strategies to decrease the high volume of applicants per residency program position so that programs can spend more time on a holistic review of potential candidates.

Unintended consequences of a broader approach to assessment include most notably the consequences of pass/fail grading in core clerkships that must be considered. Opponents will fear that United States Medical Licensing Examination Step 1 scores will carry higher weight. However, residency selection committees who focus solely on this metric around medical knowledge and test taking would shortchange their needs and bypass relevant information about holistic aspects of clinical performance. Residency programs selecting applicants will fear the risk of voluminous performance descriptions in place of grades and seek concise summaries of performance across competencies. Others may fear a surge in away rotations among students seeking grades for their transcript, but because these occur in the fourth year, students could just as readily earn grades in similar clerkships at their home institutions.

Conclusion

Clerkship grading causes multiple unintended consequences for learners with implications for their eventual independent practice. Grading generates stress and confusion for early learners, and that can interfere with desired approaches to learning and disadvantage students in ways that exacerbate inequities. Eliminating grades in favor of

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.

competency-based assessment systems focused on assessment for learning holds promise to engage learners in the process of developing essential patient care and teamwork skills and to foster their development of lifelong learning habits.

Funding/Support: University of California, San Francisco School of Medicine receives funding from the American Medical Association Accelerating Change in Medical Education program.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

Previous presentations: K.E. Hauer presented this work at the American Medical Association Accelerating Change in Medical Education spring consortium meeting, April 10, 2018, Providence, Rhode Island.

K.E. Hauer is associate dean for assessment and professor, Department of Medicine, University of California, San Francisco, San Francisco, California; ORCID: https://orcid.org/0000-0002-8812-4045.

C.R. Lucey is vice dean for education and professor, Department of Medicine, University of California, San Francisco, San Francisco, California.

References

1 Spring L, Robillard D, Gehlbach L, Simas TA. Impact of pass/fail grading on medical students' well-being and academic outcomes. Med Educ. 2011;45:867–877.
2 Green M, Jones P, Thomas JX Jr. Selection criteria for residency: Results of a national program directors survey. Acad Med. 2009;84:362–367.
3 Osman NY, Walling JL, Mitchell VG, Alexander EK. Length of attending–student and resident–student interactions in the inpatient medicine clerkship. Teach Learn Med. 2015;27:130–137.
4 Howley LD, Wilson WG. Direct observation of students during clerkship rotations: A multiyear descriptive study. Acad Med. 2004;79:276–280.
5 Feldman W, Dhaliwal G. When professors don't return essays. Teach Learn Med. 2015;27:431–434.
6 Bernabeo EC, Holtman MC, Ginsburg S, Rosenbaum JR, Holmboe ES. Lost in transition: The experience and impact of frequent changes in the inpatient learning environment. Acad Med. 2011;86:591–598.
7 Gingerich A, Regehr G, Eva KW. Rater-based assessments as social judgments: Rethinking the etiology of rater errors. Acad Med. 2011;86(10 suppl):S1–S7.
8 Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: Faculty interpretations of narrative evaluation comments. Med Educ. 2015;49:296–306.
9 Teherani A, Hauer KE, Fernandez A, King TE, Lucey CR. The relationship between underrepresented in medicine student status and clinical performance, grades and recognition: The amplification cascade with serious consequences. Acad Med. In press.
10 Krupat E, Borges NJ, Brower RD, et al. The Educational Climate Inventory: Measuring students' perceptions of the preclerkship and clerkship settings. Acad Med. 2017;92:1757–1764.
11 van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. Med Teach. 2012;34(3):205–214. doi:10.3109/0142159X.2012.652239
12 Boud D. Feedback: Ensuring that it leads to enhanced learning. Clin Teach. 2015;12:3–7.
13 Powell DE, Carraccio C, Aschenbrener CA. Pediatrics redesign project: A pilot implementing competency-based education across the continuum. Acad Med. 2011;86:e13.
14 Hauer KE, Hirsh D, Ma I, et al. The role of role: Learning in longitudinal integrated and traditional block clerkships. Med Educ. 2012;46:698–710.
15 Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. Med Educ. 2013;47:362–374.

Copyright © by the Association of American Medical Colleges. Unauthorized reproduction of this article is prohibited.