



Grade Inflation in the Family Medicine Clerkship

Kimberly Zoberi Schiel, MD; Kelly M. Everard, PhD

BACKGROUND AND OBJECTIVES: Medical educators perceive grade inflation to be a serious problem. There is some literature discussing the magnitude of the problem and ways to remediate it, but little literature is available in the field of family medicine. We sought to examine what methods of remediating grade inflation have been tried by family medicine clerkship directors, and what factors influence the chosen method of addressing this problem.

METHODS: We conducted a national Council of Academic Family Medicine's (CAFAM) Educational Research Alliance (CERA) survey of family medicine clerkship directors, inquiring about their perceptions of the seriousness of grade inflation, whether it was perceived as a remediable problem, and what methods had been tried within the last 3 years to address this problem.

RESULTS: The response rate was 69%. Clerkship directors' perceptions that grade inflation is a serious problem either nationally or in their own clerkship did not correlate with how they weighted the objective versus subjective portions of the clerkship grade. Clerkship directors who agreed that grade inflation was a remediable problem had a higher percentage of nonexamination objective criteria and a lower percentage of subjective criteria in their grading formula. Clerkship directors who agreed grade inflation is a problem in their clerkship were more likely to have tried giving feedback to graders on grade distribution than those who didn't think grade inflation was a problem.

CONCLUSIONS: Family medicine clerkship directors perceive grade inflation to be a serious problem, both at a national level and in their clerkships. Various methods of addressing grade inflation have been tried by family medicine clerkship directors.

(Fam Med. 2019;51(10):806-810.)

doi: 10.22454/FamMed.2019.614620

Grade inflation is a topic frequently discussed among family medicine clerkship directors, as evidenced by the frequency of this topic being presented at past Society of Teachers of Family Medicine (STFM) Conferences on Medical Student Education and posted on STFM discussion boards. However, there is a dearth of published literature specific to the family medicine domain. Concerns about the

effects of grade inflation on the validity of assessment of medical students appear to be well founded. A national survey found dramatic variation in grading criteria, grading terminology, and grade distribution across all clerkships.¹ This variability was found between schools and even between clerkships at the same school. Grade inflation has been well documented in internal medicine,² where 78% of clerkship directors perceived

grade inflation as a serious or somewhat serious problem, and 38% reported having given a passing grade to a student who should have failed the clerkship. The high frequency of grade inflation raises concerns about patient safety, inasmuch as students who receive high grades may have a false sense of their own competence.³ Although the problem likely exists in family medicine, there is little timely literature available.

Many reasons are cited for grade inflation: clinical grading is inherently subjective, clerkship directors may wish to avoid having to deal with upset students, and clerkship directors want to help students get into the best possible residencies.² One proposed solution is using multiple assessment tools to contribute to the final grade. When attempted in a psychiatry clerkship, this produced a more divergent spread of assigned clerkship grades.⁴ In a neurology clerkship, adding a semiobjective bedside examination evaluation to the usual subjective evaluation form resulted in a lower mean for the final grade.⁵ One institution tried breaking the clerkship grade into two components: an exam-determined grade and a faculty-determined clinical grade. The two grades were poorly correlated, as the clinical grade was higher than the written exam grade 98% of the time.

From the Saint Louis University School of Medicine, St Louis, MO.

The written exam grade also had a more normal distribution.³ Another study demonstrated that a shift from the use of nominal categories such as honors, high pass, and pass to a semiquantified system of “top 5%, top 25%, as expected, below expected and far below expected” did succeed in producing a more normal distribution of student scores.⁶ Lastly, some authors have suggested using a predetermined percentage to determine how grades are distributed.² For example, 25% of students receive honors, 25% high pass, and 50% pass.

To examine grade inflation in the family medicine clerkship, we described family medicine clerkship directors attitudes toward this topic. We further described whether those who perceived a problem tried any interventions to remediate, or the grade inflation and what specific educational interventions were tried. In this context, we use “remediate” to mean a long-term solution, not a short-term correction. We hypothesized that clerkship directors who perceived grade inflation to be a problem and who thought it remediable would give higher weights to objective criteria, lower weights to subjective criteria, and would have tried interventions to remediate grade inflation problems.

Methods

Survey

Data were gathered and analyzed as part of the 2018 Council of Academic Family Medicine’s (CAFM) Educational Research Alliance (CERA) survey of family medicine clerkship directors.⁷ CAFM members were invited to propose survey questions for inclusion into the CERA survey. Approved projects were assigned a CERA research mentor to help refine questions. The final draft of survey questions were then modified following pilot testing.

The survey was emailed to 128 US and 16 Canadian family medicine clerkship directors between June and August 2018. Invitations to participate in the study included

a personalized greeting and a letter signed by the presidents of each of the four sponsoring organizations with a link to the survey, which was conducted through the online program SurveyMonkey. Reminders were sent to nonrespondents weekly for 5 weeks. A final request was sent 2 days before closing the survey. Additionally, clerkship directors were contacted through personal email to verify their status as clerkship directors, check accuracy of email addresses, and encourage participation. The American Academy of Family Physicians Institutional Review Board approved the study in June 2018 and data were deidentified prior to the authors receiving them for analysis.

Survey Questions

The survey used standard demographic questions to determine characteristics of the clerkship directors and their clerkships. Participants answered questions about whether grade inflation was a serious problem in their clerkship or at a national level and whether grade inflation was remediable (1 to 5 Likert scale, where 1 was strongly agree and 5 was strongly disagree). Before analyses, responses were reverse-coded so higher scores were positive, thus 1 was strongly disagree and 5 was strongly agree). Clerkship directors also answered questions regarding grade inflation interventions they had tried. Finally, they indicated the percentage of the final clerkship grade that was determined by the shelf or other standardized exam, other objective criteria such as an objective structured clinical examination (OSCE) or quiz, semiobjective criteria such as history and physical examination (H&Ps) or SOAP notes, and subjective criteria such as preceptor evaluations or presentations. These are shown in Table 2.

Analyses

We summarized demographic variables, interventions to remediate grade inflation, and final grade determination using frequencies.

Independent samples *t* tests determined if agreeing that grade inflation was a problem or if grade inflation was remediable were associated with differing proportions of the clerkship final grade: shelf or other exam, other objective criteria, semiobjective criteria, and subjective criteria. χ^2 analyses were conducted to determine the association between the perception that grade inflation was remediable and whether remediation interventions were tried.

Results

A total of 99 out of 144 clerkship directors (69%) responded to the survey. Three respondents did not complete the survey and were not included in the analyses. Most of the clerkship directors were female (66%), white (78%), and had an average of 29% protected time as clerkship director. Most clerkships (73%) were block only and were either 6 (41%) or 4 (31%) weeks long. For the questions about grade inflation, responses were dichotomized where strongly agree and agree were combined to create an agree category, and the neutral, disagree, and strongly disagree responses were combined to create a do not agree category (Table 1). Neutral responses were placed into this category because they were thought to be likely to produce similar actions or lack of action as those in the disagree categories. Participants indicated which of a list of possible interventions they tried in the last 3 years to remediate grade inflation (Table 2).

Analysis of the weighting of the various components of the final clerkship grades revealed the following: subjective criteria such as preceptor evaluations or case presentation accounted for 50.4% of the final grade. Objective criteria including shelf or final exam accounted for 27.7%, other objective criteria such as OSCE or quizzes accounted for 12.8%, and semiobjective criteria such as SOAP notes accounted for 9.2%. Independent samples *t* tests showed that clerkship directors who agreed that grade inflation was a

Table 1: Clerkship Directors' Perception of Grade Inflation

Perception Statement	% Agreed
Grade inflation is a serious problem in your family medicine clerkship.	54.3
Grade inflation is a remediable problem in your family medicine clerkship.	45.7
Grade inflation is a serious problem at a national level.	73.1

Table 2: Percentage of Clerkship Directors Who Tried Interventions to Remediate Grade Inflation in the Last 3 Years

Intervention	% Who Tried
Giving feedback to graders on desired distribution of grades	49.5%
Increasing the number of assessment tools that contribute to the grade	33.3%
Breaking the overall clerkship grade into exam grade and clinical grade components	28.3%
Increasing the percentage of the grade given to objective criteria	24.2%
Using a predetermined percentage to determine how grades are distributed (example: 30% honors, 30% near honors, 40% pass)	24.2%
Adding graded semiobjective assignments (H&P write ups, SOAP notes)	22.2%
Nothing was tried	16.2%
Adding a graded OSCE	15.2%
Ranking students by percentile rather than granting a nominal grade like honors, near honors, or pass	6.1%

problem in their own clerkship or at the national level did not weight objective and subjective criteria differently from those clerkship directors who did not agree that grade inflation is a problem. However, clerkship directors who agreed that grade inflation was remediable had a higher percentage of objective criteria other than an exam (eg, OSCE or quiz) and a lower percentage of subjective criteria (preceptor evaluations or case-based learning discussions) contribute to the clerkship grade (Table 3). We ran these same analyses with only those clerkship directors who thought grade inflation was a problem and had the same results. χ^2 analyses showed that clerkship directors who agreed grade inflation was a problem in their clerkship were more likely to have tried giving feedback to graders on grade distribution than those who didn't think grade inflation was a problem (68.6% vs 32.6%, $P < .001$). No other interventions were more likely to have been tried by clerkship directors who agreed grade inflation was a problem in their clerkship.

Discussion

The majority (54%) of clerkship directors thought that grade inflation was a serious problem in their clerkship, and even more (71%) thought that it was a problem at a national level. The weighting of the objective, semiobjective, and subjective components of the clerkship grade was no different between those clerkship directors who perceived grade inflation to be a serious problem at their institution and those who did not. However, there is evidence from the nursing literature that the addition of objective criteria (multiple-choice tests) and explicit grading criteria where they were not previously in place does reduce the mean assigned grade.⁸ Our descriptive statistics show that family medicine clerkships already dedicate approximately 50% of the weighted grade criteria to objective or semiobjective elements.

In contrast, clerkship directors who perceived grade inflation to be a remediable problem in their clerkships were more likely to give a higher weighting to objective criteria such as an OSCE or quiz, and a lower weighting to subjective criteria (clinical evaluation.) The percent of weighting devoted to the exam did

not change with the clerkship director's perception of whether this was a remediable problem. It is possible that the weighting of the exam was determined at a school-wide level. It is also possible that a perception that the clerkship director had some influence over the distribution of grades led to taking action. There is evidence that the Milestones-based evaluation systems used in residencies—which are designed to be objective descriptions of competencies to be achieved—does result in better discrimination of skill level as residents progress.⁹

Clerkship directors who perceived grade inflation to be a problem in their institution were more likely to attempt the solution of giving feedback to graders on desired distribution of grades. Attempts to influence graders' patterns have been described previously. One attempt described publishing various graders' mean evaluation scores, but resulted in a paradoxical increase in mean grade given. Faculty may have wanted to avoid perception as a "hard grader."¹⁰

Believing that grade inflation is a fixable problem appears to lead clerkship directors to take action,

Table 3: Grade Determination Comparing Clerkship Directors Who Agree That Grade Inflation Is a Problem vs Those Who Don't Agree It Is a Problem

	"Grade inflation is a serious problem in your family medicine clerkship."		
	Agree (N=51)	Do Not Agree (N=43)	
% Final Grade Determined per Component	% of Grade (SD)	% of Grade (SD)	P Value
Shelf or exam	29.5 (11.8)	25.6 (14.2)	.148
Other objective criteria	12.7 (10.8)	12.6 (11.1)	.955
Semiobjective criteria	8.7 (8.6)	9.3 (10.5)	.763
Subjective criteria	49.1 (15.2)	52.5 (19.5)	.340
	"Grade inflation is a remediable problem in your family medicine clerkship."		
	Agree (N=43)	Do Not Agree (N=51)	
% Final Grade Determined per Component	% of Grade (SD)	% of Grade (SD)	P Value
Shelf or exam	28.6 (11.2)	26.9 (14.5)	.540
Other objective criteria	16.4 (10.3)	9.5 (10.3)	.002*
Semiobjective criteria	10.3 (10.7)	7.9 (8.4)	.243
Subjective criteria	44.7 (15.2)	55.7 (17.5)	.002*
	"Grade inflation is a serious problem at a national level."		
	Agree (N=67)	Do Not Agree (N=27)	
% Final Grade Determined per Component	% of Grade (SD)	% of Grade (SD)	P Value
Shelf or exam	27.8 (12.0)	27.4 (15.5)	.890
Other objective criteria	11.8 (10.2)	14.7 (12.3)	.233
Semiobjective criteria	9.0 (8.4)	8.9 (12.1)	.962
Subjective criteria	51.4 (16.7)	48.9 (18.9)	.538

*Significant at .004, Bonferroni adjustment of *P* value to account for multiple comparisons.

but it is not clear which actions will lead to effective outcomes. The attempted interventions were varied, from changing the weighting of various components of the grade, to giving feedback to graders, to adding curriculum that was more objective in nature. Clerkship directors who thought that grade inflation was a fixable problem were less likely to have tried nothing (ie, more likely to have tried something) in the last 3 years. Although no one method of trying to fix the problem was adopted by a majority of clerkship directors, the majority tried something. Giving feedback on the desired distribution of grades to graders was the most commonly tried method. This high frequency of attempting some intervention speaks to the

urgency of research toward effective methods of achieving a realistic and fair distribution of grades.

In medical education grade inflation is commonly understood to mean a high number of students receiving the top ranking.² Given that grades are one of the main criteria used by residencies to distinguish medical students from one another (ie, to rank them), a distribution of grades across the categories is necessary for any ranking to be meaningful. In contrast, competencies or milestones reflect the achievement of a certain body of knowledge or set of skills. Currently, some medical educators are calling for more competency-based criteria, in the form of entrustable professional activities.¹¹ Whether core clerkship

grades should be used to rank students or to reflect a stepwise growth in competency level is an area of active discussion.¹²

Clerkship directors are encouraged to consider their goals in grading students. If student grades are to be used as a method of comparing students to one another, then a prescribed distribution is necessary for such comparisons to be meaningful. In this case, clerkship directors are encouraged to examine the distribution of grades within their clerkships and to consider how to make their grade distribution congruent with the stated goal distribution. If, on the other hand, grades are purely meant to reflect an achieved competency, then grades may not be a useful tool for such comparison.

As with all surveys, the data are only as good as the questions asked. We did not define the term “grade inflation.” We intended this term to be understood as the phenomenon of assigning a large number of high grades (honors) within a given cohort of students. We believe that this is the common use of this term in medical education literature. However, it is possible that some respondents may have understood the term to mean the rise of the average grade for a cohort over time. The questions required clerkship directors to reflect on a possible weakness in their own clerkships, which is never an easy task. This may account for the higher perception of grade inflation being a problem at a national level than in the respondent’s own clerkship. In addition, it is possible that there was some ambiguity regarding the perception of grade inflation being remediable. This response implies that a problem can be fixed but has not yet been fixed in the respondent’s clerkship. Another limitation was the failure to acknowledge that a small percentage of schools grade their clerkships on a pass/fail system.¹ Lastly, we only inquired about solutions tried within the last 3 years. It

is possible that some clerkship directors tried the listed solutions more than 3 years ago, or tried other solutions. We limited the time frame to 3 years because we were interested in a current problem and recently-tried solutions. Some data may have been missed by this time limitation, but the data that were captured imply an important, pressing problem without a clear solution.

CORRESPONDING AUTHOR: Address correspondence to Dr Kimberly Zoberi Schiel, Family and Community Medicine, 1402 S Grand Blvd, St Louis, MO 63104. 314-977-8480. Fax: 314-977-5268. kimberly.zoberi@health.slu.edu.

References

- Alexander EK, Osman NY, Walling JL, Mitchell VG. Variation and imprecision of clerkship grading in U.S. medical schools. *Acad Med.* 2012;87(8):1070-1076.
- Fazio SB, Papp KK, Torre DM, Defer TM. Grade inflation in the internal medicine clerkship: a national survey. *Teach Learn Med.* 2013;25(1):71-76.
- Paskausky AL, Simonelli MC. Measuring grade inflation: a clinical grade discrepancy score. *Nurse Educ Pract.* 2014;14(4):374-379.
- Roman BJ, Trevino J. An approach to address grade inflation in a psychiatry clerkship. *Acad Psychiatry.* 2006;30(2):110-115.
- Schmahmann JD, Neal M, MacMore J. Evaluation of the assessment and grading of medical students on a neurology clerkship. *Neurology.* 2008;70(9):706-712.
- Weaver CS, Humbert AJ, Besinger BR, Graber JA, Brizendine EJ. A more explicit grading scale decreases grade inflation in a clinical clerkship. *Acad Emerg Med.* 2007;14(3):283-286.
- Mainous AG III, Seehusen D, Shokar N. CAFM Educational Research Alliance (CERA) 2011 Residency Director survey: background, methods, and respondent characteristics. *Fam Med.* 2012;44(10):691-693.
- White KA, Heitzler ET. Effect of Increased Evaluation Objectivity on Grade Inflation: Precise Grading Rubrics and Rigorously Developed Tests. *Nurse Educ.* 2018;43(2):73-77.
- Kuo LE, Hoffman RL, Morris JB, et al. A Milestone-Based Evaluation System-The Cure for Grade Inflation? *J Surg Educ.* 2015;72(6):e218-e225.
- Schell SR. Displaying faculty grade averages during 3rd year medical student clerkship evaluations: effects upon grade inflation. *J Surg Res.* 2003;114(2):260.
- Lomis K, Amiel JM, Ryan MS, et al; AAMC Core EPAs for Entering Residency Pilot Team. Implementing an Entrustable Professional Activities Framework in Undergraduate Medical Education: Early Lessons From the AAMC Core Entrustable Professional Activities for Entering Residency Pilot. *Acad Med.* 2017;92(6):765-770.
- Hauer KE, Lucey CR. Core clerkship grading: the illusion of objectivity. *Acad Med.* 2019;94(4):469-472.