

Reflections on Fair Grades

DARYL CLOSE

There are three general issues that I raise in “Fair Grades”¹ that I want to reexamine in light of commentators’ critiques in this volume and elsewhere: whether learning always trumps Principle 2, how Principle 1 fares in the conflict between inter-instructor grading standards and academic freedom, and punitive grading.

First, in “Fair Grades,” I propose a *prima facie* necessary condition for a grade to be fair, viz., Principle 2: “Grading should be based on the student’s competence in the academic content of the course.” I argue that Principle 2 is the best of three common conceptions of grading. Both Jennifer McCrickerd and John Immerwahr object to Principle 2, holding that there are cases when fairness in grading should yield to practices that increase the probability of student learning.

Second, in “Fair Grades,” I do not explore the issue of inter-instructor grading, interdepartmental grading, and inter-institutional grading. Stephen Finn² states that my Principle 1—“Grading should be impartial and consistent”—entails such consistency beyond grading students in a specific course in a consistent manner. Finn uses the West Point policy to argue that such consistency may infringe on the academic freedom of instructors. This conclusion could therefore require weakening Principle 1; a critical outcome since most instructors regard Principle 1 as the central component of fairness in grading.

Third, the general issue of punitive grading—a species of grading on deportment—seems complicated to me. I remain committed to my earlier position that most types of punitive grading violate Principle 2, for example, grading on attendance, tardiness, sleeping in class, the worst cases of academic dishonesty, etc. However, I think that less serious cases of academic dishonesty could be treated as genuine exceptions to Principle 2.

I. Does Fairness in Grading Ever Yield to Raising the Probability of Learning?

One way to object to Principle 2—an objection that remains a central concern to me—is to argue that fairness in grading must always, or at least sometimes yield to whatever pedagogical techniques might increase the probability of learning.³ McCrickerd’s alternate definition of fairness in grading is clear: “Whether a particular grading practice is just or not is determined by whether its adoption enhances or detracts from learning.”⁴ I read this as a rejection of Principle 2. That is, fairness in grading as I regard it may accidentally coexist with good learning pedagogies, but if the two come into conflict, fairness must bow to whatever enhances learning.

On the other hand, Immerwahr’s concept of “motivational grading”⁵ is more modest—he thinks that Principle 2 is too strong because there are exceptions to the rule. I, too, think that there may be exceptions, but the bar is higher for me. As in “Fair Grades,” I still don’t rule out the possibility that a *prima facie* unfair grading practice may be acceptable; for example, in cases where we have a well-supported causal model—not just a correlation—showing that the grading practice is a probability-of-learning raiser for every student. Punitive grading may also be an exception, as I will discuss below.

What, then, can we make of this conflict between learning and fairness? Dedicated teachers think first about their students’ learning; everything else in the classroom seems secondary. But, as soon as we say this, as philosophers we immediately construct counterexamples. “Everything else is secondary to learning? Really?” If valid research showed that random student canings improved learning, would we all go cane shopping? Not likely.

So, we already know the answer to the question, Does learning trump everything else in the classroom? The answer is, “No, it doesn’t.” However, we need to consider the more specific question of whether the promotion of learning is sufficient to morally justify a particular grading practice. We are clearly in the land of line-drawing.

Some lines are easy to draw. For example, there is emerging research showing that low doses of psychostimulants such as amphetamine may increase learning.⁶ But, no instructor is likely to raise a student’s course grade because the student’s medically prescribed use of amphetamines is a learning probability raiser. Such a grade wouldn’t satisfy Principle 2 because it would have no necessary connection to the student’s knowledge of course content. McCrickerd’s alternate standard for fair grades—“Whether a particular grading practice is just or not is determined by whether its adoption enhances or detracts from learning”—is thus too strong. . Later in her comments she states that promoting learning cannot go beyond the “legitimate domain of the teacher.” She says that “the purview of the professor can be understood to cover, legitimately, only the students’ engagement with the course material regardless of the fact that learning is influenced by factors beyond this.”⁷ McCrickerd is surely right here.

Whether McCrickerd's modified standard is strong enough to justify her inclusion in course grades of such non-course-content-specific behaviors as effort, teamwork, and risk-taking⁸ is unclear, because species of these behaviors can be viewed as content-based. First, effort, per se, cannot plausibly be a grade component, since *effective* effort is what we value. And, we happen to have a very reasonable proxy for effective effort, viz., demonstrating a high level of knowledge of course content and skills. Likewise, teamwork is no stranger to the classroom and in my experience is invariably paired with content-relevant activities. Last, risk-taking, per se, does not seem to be a likely candidate for a grade component. What we value is *reasonable, content-relevant* risk-taking such as participating in the discussion of a controversial issue in an applied ethics course or a philosophy of religion course. So, I see nothing here that necessarily conflicts with Principle 2.⁹

What about Immerwahr's concept of "motivational" grading? Is it an exception to Principle 2? In an earlier paper, Immerwahr defines motivational grades as "those that are given solely or primarily for the purpose of encouraging behaviors that are likely to improve learning, such as attending class having done the assigned reading for that day."¹⁰ Immerwahr's example is interesting since I never argue that grading class preparation is unfair. Rather, I argue against grading on attendance, a view with which Immerwahr agrees.¹¹

Immerwahr's discussion of motivation grading is carefully qualified; in fact, so qualified that it is not easy for the reader to find cases of common grading practices that are both motivational and that meet his standards of acceptable motivational grading. The closest cases seem to be in course designs that use so-called "active learning" pedagogies. Immerwahr gives the example of blog posts in which students are graded not only on the content of their posts, but also whether they have made the requisite number of posts. It is this latter grade component that he views as being employed motivationally.

On this account of motivational grading, it seems that one is grading motivationally just in case a grade is assigned to a learning activity in which the student meets or fails to meet some learning objective that doesn't require *measuring* course content knowledge. This would include not just counting blog entries, but also tabulating class discussion occurrences, participating in a group debate, turning in three reflection papers instead of five, and the like. Most of us use grading schemes such as these. I began using message boards ("forums") more than 20 years ago. I read them for content and students who miss a post on a given topic get a zero for that topic. Have I been using motivational grading all of this time?

In my view, these learning activities are easily covered by my Principle 2.1: "Grades should be assigned on the basis of an expert evaluation of student work."¹² Such activities constitute student work. Principle 2.1 excludes grading on comportment in the general sense, unless it is inherently relevant to course content. We may choose to refer to grading on the basis of "credit for engaging with the

course material” as “motivational.” However, it seems to lack the core of what I would regard as motivational grading, viz., where students are told that they will be marked up or down on the basis of behaviors that have nothing to do with their knowledge of course content, e.g., tardiness, number of absences, playing with a cellphone in class, being cheerful or sullen, sleeping in class, buying the required textbooks, etc. As learning probability raisers/reducers, these examples certainly fit under Immerwahr’s definition of motivational grading, but I’m not sure that he would regard all of them as generally acceptable components of a course grade. As with McCrickerd, I think that Immerwahr and I are not far apart on the reach of Principle 2. They both may object to my rapprochements, of course.

II. Inter-Instructor Consistency in Grading

Using a common syllabus and grading rubric for large, multi-section courses like those reported by Finn at West Point is a long-time practice in the academy. For the issue of fairness in grading, the common-syllabus, collaborative-grading approach undercuts instructor shopping, that well-established “beauty” contest in which students design their schedules to produce a semester that yields the highest grades for the least amount of work. The collaborative grading system at West Point is almost certain to increase fairness in grading for that reason alone. Such practices take Principle 1—“Grading should be impartial and consistent”—to the next level beyond the individual instructor.

Cooperative course designs and grading are benign if they are the result of true collaboration among peers, rather than administratively mandated. However, as Finn notes, requiring instructors to follow a syllabus other than their own is an infringement of the academic freedom of the instructor. So, here, one core academic value, fairness in grading via Principle 1, is in direct conflict with another core value, academic freedom. The dilemma can be summarized like this: Principle 1 requires consistency in grading; consistency in grading requires standardization of grading among instructors of multiple-section courses; standardization requires faculty to surrender a significant part of their academic freedom as professionals to determine what they deem best for their students.

The dilemma can be resolved by modifying Principle 1 so that impartiality and consistency in grading cannot tread on academic freedom. This approach would require that inter-instructor coordination of course design and grading be the result of an explicit and voluntary agreement among instructors, *as peers*, and independent of any administrative policy. By Finn’s description, that’s not the way it works at West Point, nor, I would wager, at other institutions that use common syllabi, exams, and grading practices in multi-section courses. I would be surprised to learn that at such institutions the social contract gets recreated every time a new faculty

member arrives in the department. It is more likely to be a case of, “You’re going to be teaching a section of Introduction to Philosophy, and here’s how it works . . .”

There is a related problem with standardizing course designs and grading across multiple sections, viz., the “unbundling” of faculty work associated with distance education and for-profit institutions that rely on a high fraction of adjunct instructors. Unbundling applies the division of labor to faculty work, breaking our work into its many constituents such as program design, course design, syllabus construction, content “presentation,” grading, faculty-student interaction (office hours), and so on. When accompanied by attacks on the system of tenure and the replacement of tenure lines with poorly paid adjuncts, this phenomenon is far from benign.

For the present discussion, we need only note that the unbundling of faculty work—and the consequent industrialization of the academy and loss of academic freedom—is not the same thing as standardized course design and grading. In the distance education market, they may arise from the same motivation, viz., cost reduction, but they are logically distinct. Faculty can work collaboratively, developing specialties in the various parts of academic work, without necessarily sacrificing their academic freedom or that of their colleagues.

The tricky bit is to maintain faculty control of the unbundling process so that it remains benign. The fact that such control is fiercely resisted by regents and governing boards who subscribe to the “students-are-customers-and-faculty-are-factors-of-production” model of higher education shouldn’t lead us to believe that faculty can’t exert control. It does mean that faculty—especially tenured faculty—must be much more politically active, particularly at the local level. While at-will faculty on term contracts are probably not able to aggressively challenge institutional efforts to unbundle online courses, they can decline participation in the outsourcing of grading unless they are grading students with whom they have an instructional relationship. All faculty, tenured or not, can assert their intellectual property rights in all of their instructional documents, lecture podcasts, LMS Web sites, and so on. This alone makes it more difficult for the unbundling of faculty work to pass without license from faculty control to institutional control. Academic policy committees can write policy that forbids the assignment of a grade by any person other than the faculty member who instructed the student and evaluated the student’s work. That prevents “dummy” instructors-of-record who submit course grades to the registrar on behalf of the actual instructors, e.g., independent contractors who have no faculty status but perform various instructional subtasks under the alleged “direction” of the instructor-of-record—TAs excepted, of course.

III. Punitive Grading

While my views regarding the most typical sorts of punitive grading, attendance, tardiness, and the like, remain unchanged, there is a distinction in cases of punitive grading that I did not draw in “Fair Grades.” The distinction concerns grade penalties that are unconnected with course grade components.

Case 1: Assume that the attendance grade component in my Logic course forms 10 percent of the course grade. Suppose my policy states that after eight cuts, the student will receive a zero on that grade component. This means that the course grade will be reduced by one full letter grade on a 60-pass grading scale.¹³ Students who do not exceed eight cuts automatically achieve a score of 100 percent on the attendance grade component. If I reduce Fred’s course grade because he has exceeded eight cuts, this seems to be punitive grading—excepting performance content courses such as theatre, music, etc. I am punishing the student for missing class by reducing the course grade on grounds other than the student’s knowledge of course content and skills. By Principle 2, this is an unfair grade.

Case 2: Suppose that my academic honesty policy in Logic states that obtaining help from another student or accessing a source of information—e.g., a crib sheet—during the closed book final exam will result in the course grade being reduced by one full letter grade—again 10 percent.¹⁴ I catch Sally using her cellphone during the final to look up material in the textbook and lecture notes, so the grade penalty operates and her C in the course becomes a D.

The cases differ because in Case 2 the grade penalties are external to the ordinary grade components of the course. Put another way, unlike the attendance penalty, there is no course component called “Didn’t cheat on the final exam” for which non-cheating students automatically achieve a score of 100 percent of that grade component. Rather, in Case 2, I am punishing the student for a serious violation of academic honesty, not for failing to meet an attendance grade component.

While both cases are instances of grading on comportment, I think that only Case 1 is clearly unfair.¹⁵ As I suggested in “Fair Grades,” punitive grading in cases of academic dishonesty may be genuine exceptions to Principle 2, i.e., they are not unfair despite the fact that the course grade will no longer be a function of the instructor’s best estimate of the student’s knowledge of course content and skills.

In the case of egregious violations of academic honesty such as buying a term paper or using an earpiece and a confederate during a high-stakes exam, my preference would be something like immediate expulsion and administrative withdrawal from all courses. I still find this better than punitive grading. However, Immerwahr makes a strong case for punitive grading in some academic dishonesty cases, noting that “employers and graduate schools might feel more lied to if they learned that academically dishonest students receive As in courses where they participated in cheating.”¹⁶ Here, Immerwahr is referring to my example in which the instructor drops an A student to a C because she allowed a friend to copy from her exam. In

such cases, a punitive grade is a lesser penalty than expulsion, so the practice perhaps could be defended on grounds of proportionality. An alternative to a punitive grade would be to handle those cases through the Student Affairs judicial system that punishes violations of the student code with eviction from student housing, expulsion from fraternities or sororities, immediate suspension from intercollegiate athletics, and so on.

The case of punitive grading for late papers is also difficult. First, Immerwahr states that it is “unjust to those who hand their papers in on time not to punish the late paper.”¹⁷ This is plausible in some cases and implausible in others. I think our intuitions here reflect our moral assessment of the reason for lateness. If Fred’s paper misses the Friday deadline because the big weekend party started Thursday night when he should have been completing his paper, the late penalty is a grade for his weakness of will, his intellectual immaturity, and the like. If his paper is late because he spent every waking moment over the past week caring for a sick child, a late penalty is a slap in the face; a punishment for being a good parent. If his paper is accepted late without penalty because his track meet away provides him with a University-approved excuse, wouldn’t the on-time student have a valid protest? On the one hand, we want to say, “Your paper is late. It doesn’t matter why,” while at the same time we are inclined to draw fine, even arbitrary, distinctions among the reasons for being late. In the former case, we’re being obtuse; in the latter, we’re grading on moral virtues.

Second, Immerwahr states that “not penalizing late papers would be wildly impractical,”¹⁸ but there are alternatives. I agree with Immerwahr’s point that the late paper should (sometimes) be treated differently than the on-time paper, but it doesn’t follow that the late paper must be penalized. One alternative that I use is an “on-time bonus.” This allows me to mark papers on their merits whether on-time, or late, it preserves the deadline, and it avoids “penalizing” the on-time student by treating her no differently than the lazy student. One might object that this is a difference in name only, but the two approaches are quite different; one is punishment, the other is reward. The real problem with the on-time bonus system is that in some courses, it will amount to grade inflation. That’s not unfair as long as it is available to all, but some may find it no less objectionable than punitive grading for lateness because it appears to violate Principle 2. This is not to trivialize grade inflation, of course, but possibly inflating a grade is a far cry from a punitive reduction, so it is at least in the spirit of the principle if not the letter.

IV. Conclusion

There are several topics that deserve more thought. First, grading on class rank, i.e., grading that is based on relative ranking the students in the course, especially by forcing the scores into a normal distribution or by some other peer-relative

ranking system still seems very wrong to me. The idea that we know in advance of examining student work in a given course that some students will receive a high course grade and others will fail the course is unempirical; silly, in fact. Grading on the normal curve is unfair because it can deny a high-achieving student a correspondingly high course grade merely because her average sits at the bottom of the normal curve. My point here concerns the relatively small data sets that occur in a specific course. As I note in “Fair Grades,” I’m not militating against the central limit theorem or the law of large numbers.

Second, I applaud McCrickerd’s discussion of the pernicious belief prominent in U.S. culture that success is largely a function of innate talent.¹⁹ While I think that this issue has little to do with grading itself, it is imperative that teachers at all levels combat the talent ideology. I don’t agree with McCrickerd’s view that we should grade on effort, but we should encourage and endorse hard work among our students as the most critical means of accomplishing personal goals, educational or not.

Last, the institution of grading itself should not be regarded as a minor matter. The wide range of opinions expressed here about grading—if at all representative, as I suspect it to be—raises serious concerns about both the validity and the reliability of grading in higher education. Put simply, we need to reconsider how we evaluate students. For example, as Immerwahr observes, “badging” should be reviewed as a supplement to grades.²⁰ Another alternative to grading is some form of narrative evaluation. Hybrids of pass-fail grading and narrative evaluation have been widely adopted by both mid-tier and first-tier undergraduate institutions and professional schools, including medical schools.²¹

Notes

I want to thank the editors for their comments and suggestions. Their help improved this paper.

1. Daryl Close, “Fair Grades,” in *Philosophy Through Teaching*, ed. Emily Esch, Kevin Hermberg, and Rory E. Kraft, Jr. (Charlottesville, VA: Philosophy Documentation Center, 2014), 139–172.
2. Stephen Finn, “Fair Grades and Inter-Instructor Grading Consistency,” in *Philosophy Through Teaching*, ed. Emily Esch, Kevin Hermberg, and Rory E. Kraft, Jr. (Charlottesville, VA: Philosophy Documentation Center, 2014), 185–188.
3. We should be careful in attributing causality to pedagogical techniques—including the use of grades—in producing learning. Leading the horse to the water may be highly correlated with water drinking, but whether the former causes the latter is a different question.
4. Jennifer McCrickerd, “What Students Deserve: A Grading Policy that Supports Learning,” in *Philosophy Through Teaching*, ed. Emily Esch, Kevin Hermberg, and Rory E. Kraft, Jr. (Charlottesville, VA: Philosophy Documentation Center, 2014), 173–177, 173.

5. John Immerwahr, "The Case for Motivational Grading," *Teaching Philosophy* 34:4 (2011): 335–346.
6. For example, see M. Elizabeth Smith and Martha J. Farah, "Are Prescription Stimulants 'Smart Pills'? The Epidemiology and Cognitive Neuroscience of Prescription Stimulant Use by Normal Healthy Individuals," *Psychological Bulletin* 137:5 (2011): 717–741.
7. McCrickerd, "What Students Deserve," 176.
8. Jennifer McCrickerd, "What Can Be Fairly Factored into Final Grades," *Teaching Philosophy* 35:3 (2012): 278.
9. McCrickerd is not likely to accept my adjustments of the concepts of effort, teamwork, and risk-taking in grading. She wants to include assessments of these activities, simpliciter, in a course grade because they may be valued by employers and other transcript readers.
10. Immerwahr, "The Case for Motivational Grading," 336.
11. *Ibid.*, 344–345.
12. Close, "Fair Grades," 158.
13. When I was an undergraduate, the attendance policy for all freshmen was far more drastic: an automatic F in the course after the second cut.
14. Normally, such a policy will be paired with an automatic zero on the exam itself, so the total penalty will possibly exceed a one-letter-grade reduction, depending on the weight of the exam. The typical dishonesty case is thus a hybrid of the two examples presented here. At my institution, the Dean may impose a further penalty including an F in the course and permanent expulsion from the University.
15. For example, consider a grading policy in which the student can have three cuts, excused or unexcused, but the fourth cut results in a letter grade reduction, the fifth, another letter grade reduction, etc., until the seventh cut yields an F in the course. Such a policy is certain to motivate faithful class attendance. A student with a chronically ill child is equally certain to get a low mark in the course, or even fail the course, despite the fact that she may have mastered the course material at a very high level. Contrast this with an orchestral ensemble course in which various types of comportment form some of the course content itself, e.g., attending rehearsal, dressing in the prescribed uniform for performances, resting one's instrument in the prescribed way, etc.
16. John Immerwahr, "Grades: Definitive Picture or Just a Snapshot?," in *Philosophy Through Teaching*, ed. Emily Esch, Kevin Hermberg, and Rory E. Kraft, Jr. (Charlottesville, VA: Philosophy Documentation Center, 2014), 178–183, 182.
17. *Ibid.*, 182.
18. *Ibid.*
19. See Geoff Colvin, *Talent Is Overrated* (New York: Penguin Group, 2008).
20. For example, see Julie Jones and Nathan Altadonna, "We Don't Need No Stinkin' Badges: Examining the Social Role of Badges in the Huffington Post," *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)* (2012): 249–252.
21. For example, see University of Arizona College of Medicine Educational Policy Committee, "Program-Wide Grading and Progression Policy" (Tucson: University of Arizona College of Medicine, June 20, 2013), http://epc.medicine.arizona.edu/node/27#GradingProgressYrs3_4.