

Published in *Teaching Philosophy* 32:4 (December 2009), 361-398. Available online with POESIS license. Offprints available on request at [dclose@heidelberg.edu](mailto:dclose@heidelberg.edu). Citations should be to the published version only.

## FAIR GRADES

Daryl Close  
Heidelberg University

*Abstract:* Fair grading is modeled on two fundamental principles. The first principle is that grading should be impartial and consistent. The second principle is that a fair grade should be based on the student's competence in the academic content of the course. I derive corollary principles of fair grading from these two basic principles and use them to evaluate common grading practices. I argue that exempting students from completing certain grade components is unfair, as is grading on attendance, class rank, department, tardiness, effort, institutional values, moral virtues such as cheerfulness and helpfulness, and other non-course-content criteria.

What is fair grading? For example, is there a fair way to average a series of letter grades? Are such practices as “grading on the curve,” peer grading, and dropping the lowest quiz score from each student's course average fair? Is it fair to excuse “A” students from the final exam? Should students be graded on attendance, tardiness, or other forms of comportment? Exactly what is “extra credit” or “make-up” work? Are these practices fair? What is the difference between a “hard grader” and an “easy grader?” Is the former less fair than the latter? Certain unfairnesses must be reasonably obvious or there would not be such predictable protests from students when their teachers violate basic principles of fairness. Other principles of fair grading must certainly be less obvious, given the wide range of grading practices that teachers employ.<sup>1</sup>

Most professors learn to grade as teaching assistants in graduate school. While some graduate programs provide substantial training in teaching one's discipline, many of us learned by imitation. As a second year graduate student, I was assigned a section of Introduction to Philosophy—arguably one of the most difficult courses in the undergraduate philosophy curriculum to teach—without the least bit of guidance beyond my own experience as a student. I soon observed what I thought even then were unfair or unethical grading practices. I did my best to grade fairly, but it was hardly a topic of conversation in the graduate student lounge. As I gained teaching experience over the years, I became more concerned that my grading should reflect an understanding of basic moral concepts.

Because we are well-intentioned, conscientious professors, we assume that we grade our students fairly, regardless of the particular grading techniques that we use. While we may all be conscientious and concerned graders, it doesn't follow that we are consequently fair graders. Since grades function as a

---

<sup>1</sup> While there is a huge literature on student evaluation and assessment, including the mechanics of grading, there is much less commentary on the ethical issues of grading, beyond perhaps noting that evaluation should be “consistent, equitable, and fair” (Gullickson and the Joint Committee 2003).

form of academic currency<sup>2</sup> that students can use to gain access to other valuable economic goods such as graduation honors, scholarships, social standing, graduate admissions, etc., our grading systems should be fair and we should administer them in a fair way.<sup>3</sup> Moreover, because transcript readers rely on the accuracy of grades in communicating information about the student's mastery of course content,<sup>4</sup> instructors have a professional duty of truthfulness in grading that extends beyond the student. The overall goal of this paper is to examine such questions, and to look at how some common grading procedures are ethically problematic when fairness is taken into account.<sup>5</sup>

My primary approach will be to develop a model of what grades should be and then identify a short list of principles of fair grading. I expect some of my conclusions to be provocative. I have practiced at one time or another grading policies that I now believe to be unfair or violations of professional duties. This paper is presented as a challenge to others who, like me, are greatly concerned to treat their students fairly and to fulfill their professional duties as teachers. My conclusions should not be taken to be a list of necessary and sufficient conditions for fair grading. At best, I will establish a partial list of what I believe are necessary conditions.<sup>6</sup>

## I. Grading and Evaluation of Learning

---

<sup>2</sup> Receiving a grade is not receiving an economic good or service in which only a few at the top of the grade "economy" receive the highest quality goods, while those at the bottom receive barely enough to survive academically. Not incidentally, the silliness of such a model shows at a stroke the underlying flaw in the "education is just another product" model endorsed by some colleges and universities. Cf. Alan Goldman's (1980) view that grades "serve the academically superfluous but socially vital function of sorting people for later educational opportunities and careers."

<sup>3</sup> In this paper, I assume, rather than argue that we should grade our students fairly. By "fair" I mean to borrow on our most common and shared concepts of justice. I generally avoid substantive excursions into the theory of justice. In particular, I make no original or controversial claims about justice or desert.

<sup>4</sup> I will use "course content," "course material," and similar expressions to refer generally to the learning goals of the course, whether they are skills, knowledge, beliefs, attitudes, dispositions, etc.

<sup>5</sup> One may object to the whole premise of this paper, viz., that fairness is of paramount concern in grading. The objection is that student learning counts more than fairness, so if treating a student fairly in one's grading system gets in the way of student learning, fair grading should get out of the way. I identify this view below as an instrumental conception of grading—grading is just a tool for achieving other academic goods.

<sup>6</sup> Disclaimer: I am not concerned here with the wisdom or folly of academic grading, so nothing I argue in this paper relies on a particular justification of the social practice of academic grading, *per se*. And, while there is a long history of philosophical analysis of the general concept of evaluation that dates back to Aristotle, nothing that I advance in this paper depends on a particular model of evaluation, e.g., J. O. Urmson's instructive paper, "On Grading" (Urmson 1950).

---

It is clear that evaluating student mastery of course content is not the same as assigning a grade to a student. For one thing, evaluation of students is very old, while grades are historically very recent—in the U. S., they appear in the mid-1800s. Plato may have frequently, even continuously, evaluated his Academy students, but there is no record of Plato issuing grades. Furthermore, while we may evaluate our students without grading them, a grade may be issued to a student without any evaluation of the student’s learning. Such a grade is the *sine qua non* of professional misconduct in grading and is an intuitively unfair grade, but it is a grade nonetheless as long as it was officially recorded by an institutional agent on the student’s grade record or transcript. Viewed as a species of judgment, an evaluation can be made without a public report of the judgment, and a public report of a judgment can be made in the absence of any judgment at all—a lying report, literally.

So, even though evaluation and grading, per se, are not logically related, evaluation and fair grading are connected, viz., a fair grade must at least include a proper evaluation of student work. As I will argue below, the student work that is evaluated must be connected to the course content in some way. Moreover, the evaluation must be accurate and the accuracy presumably must arise from the application of instructor expertise. That is, it can’t be a lucky guess, or the result of a random process, nor can it be the mere dictum of a political authority in the institution. So, let’s assume as a rough definition that a “fair grade” means, at minimum, *a summary mark of an accurate, expert evaluation of student academic work* that (1) is normally made by the course instructor, (2) appraises a student’s knowledge and/or skills in the subject matter of the course, and (3) is permanently recorded in a uniform way by the instructor in the student’s institutional record.<sup>7</sup>

Grade evaluations are typically made by a single teacher, but they may be made collaboratively by more than one teacher. Nothing that I have to say about fairness is affected by this difference. Common grades are A, B, C, D, F, and P (pass), sometimes with plusses and minuses, and other distinctions. I do not treat such marks as WP (withdraw passing), WF (withdraw failing), I (incomplete), Audit, etc. as grades, per se, even though we often refer to them that way. They are marks of a purely administrative function, even though evaluation of student work may be involved. Nothing that I have to say here concerns such marks, except incidentally.

A few institutions in the U. S. evaluate the learning of their students, but do not compute grades in the usual sense of the word, providing students and third parties instead with written evaluations of knowledge of course content. Most of what I say below nonetheless applies to such institutions, showing that a fair grade is best treated as a species of evaluation of knowledge and competence in course content.

## II. Grading, Merit, and Fairness

Being graded in school is one of the few experiences of merit-based evaluation shared by virtually every formally educated adult. The work world, with its political appointments, “market” based pay scales, seniority, “dress for success,” racism, sexism, nepotism, and “old boy” favoritism, is far from being a meritarian system. Not even the justice system, with plea bargaining, unevenly distributed legal counsel, and criminal penalties that vary widely from state to state, seems to be so firmly committed to pure merit-based evaluation procedures as does academic grading.

---

<sup>7</sup> I intend both course component marks entered in the instructor’s personal gradebook as well as final course marks entered on the student’s transcript.

Consider the apocryphal professor who grades papers by throwing them down the stairs in order to rank them, A through F. Students and professors alike recognize this procedure as defective because it substitutes chance in place of merit as a grading criterion. But an equally serious flaw that often goes unnoticed is the assumption that given a stack of term papers, for example, every grade category, A through F, is sure to be represented. What evidence is there, prior to grading, that on a given exam or paper, there will be at least one F (or a least one A)? Grading “on the (normal) curve,” for example, makes just such an assumption, as does “eyeballing” a list of class scores for “natural” grade category divisions.

In such techniques, students are not graded solely on the basis of academic merit, but rather on relative class rank. Relative class rank position is demonstrably a function of class size, class composition, time of day the class is offered, and many other variables unrelated to merit. Such criteria may have statistically significant predictive power, but it does not follow that it is fair for us to invoke them, or to permit their influence, in determining students’ grades. For example, student performance on quizzes can be highly predictive statistically of performance on hour exams, midterms, and final exams. Yet, it would not be obviously fair to base a student’s course grade solely on the quiz scores, allowing the student to skip the midterm and final exams on the grounds that we can accurately predict how the student would have performed had he or she actually shown up for those exams.

A fair grade, then, is a grade that the student merits.<sup>8</sup> Since merit is frequently acquired unequally, fair grades are not necessarily equal or similar grades. This means, for example, that a faculty member’s grade distribution cannot be treated as *prima facie* evidence of fair or unfair grading. The arguments, “Professor Smith failed 70% of her students, therefore, she’s clearly an unfair grader,” and “Professor Jones gave 80% of his students an A, so Jones is obviously a very fair grader” are unsound.

Of course, access to meritorious academic achievement must be as equal as possible. But meritorious achievement itself cannot be treated as a scarce resource available only to a few. Regardless of how self-evident this point may be, its adoption is far from universal. To suppose that high academic achievement—or its recognition—is a scarce resource that must be carefully rationed is the foundation of such practices as “grading on the curve.” In such a practice, we assume that students are engaged in a competition for high marks, and since there are not “enough” high marks for everyone, only a fraction of the students—those at the top of the distribution—will receive them. So, our usual intuitions about distributive justice, i.e., the just distribution of economic (scarce) goods and services, are consequently of little use in thinking about fair grades precisely because grades (A, B, C, D, F) are not scarce resources.

---

<sup>8</sup> We can also say that a fair grade is a grade that the student deserves. Some students deserve As and some deserve Fs. But do we say that a person convicted of a misdemeanor deserves a heavy fine and a jail term, while the same action in another state is legal? Do we say that the waitress earning barely \$6 per hour “deserves” her pay, or that the CEO of General Motors “deserves” \$15 million annually in salary and benefits? The extensive literature on the concept of desert notwithstanding, what we mean by “A deserves grade X” remains refractory, except perhaps in broadly circular constructions such as “Mary deserves a high grade because she has a perfect average on all of her graded work.” Claims of punitive desert such as “Mary deserves to be expelled because she cheated on the final exam.” are generally beyond the scope of this paper.

They may be constructed as such, but they are not inherently so.<sup>9</sup> If we design our courses and our grading systems as if certain grades were a scarce good, we've already adopted an ethically questionable model of grading.

### III. Three Models of Grading

#### **Model 1: Grades are rewards and punishments for learning or failing to learn course content or institutional values.**

Model 1 holds that grades should function as a means of rewarding and punishing students across a range of academic and institutional values, including effort, amount of improvement, moral worth, promptness, discharge of civic duties, class attendance, and course content knowledge and skills. I am assuming here that the end of such rewards and punishments is restricted to student learning. On this model, the professor rewards high levels of compliance with such values with high grades, and punishes low levels of compliance with low grades. The basic idea of Model 1 is that because many students value grades more or less independently of their own learning, we can induce learning—in part—by classic conditioning, using grades. For educators, the powerful attraction of the reward/punishment approach to grading stated in Model 1 is that it is applicable across a broad variety of educational goals, both in and out of the classroom.

Unfortunately, the power of the reward/punishment model of grading is its weakness with respect to fairness. The clash between consequentialist reasoning and justice-based reasoning in ethics is an old story, but my only point here is that instrumentally grounded grading practices have gained such a foothold in the academy that they deserve our attention. Once grades are accepted as little more than the means to achieve ends such as learning course content or developing a virtuous character, issues of fairness quickly fade in importance.

For example, we know that cutting class is a good predictor of a lower course grade, all other things being equal. So, on the reward/punishment model of grading, we should design our courses to punish or reward students on the basis of their attendance behavior in order to improve their learning. But, as with any utilitarian account, achieving good in the aggregate can allow unfair treatment at the individual level. In the case of attendance, a student who learns well, but skips class frequently will be

---

<sup>9</sup> On January 21, 2005, Princeton University announced the implementation of a new policy designed to control what Princeton believes is grade inflation on its campus. The policy had been designed and approved by the Faculty the previous spring. The substantive part of the policy states

Princeton's new expectations posit a common grading standard for every academic department and program, under which A's (A+, A, A-) shall account for less than 35 percent of the grades given in undergraduate courses and less than 55 percent of the grades given in junior and senior independent work (Office of the Dean of the College 2005).

Regardless of any good end to which this policy was designed—that's a separate issue—the policy is clearly unfair in its arbitrary formulation of a grade distribution prior to the examination of any student work. Perhaps our colleagues at Princeton lacked the political will to reveal the academic “full Monty,” i.e., grading on the normal curve.

---

punished just the same as the weak student for whom attendance is essential to learning. The defender of Model 1 presumably does not see a problem with this result. But, I think that the idea of punishing a student with a low course grade for excellent performance on grade components, but who skipped class frequently, will seem counterintuitive to many teachers, despite the fact that the practice may produce a net increase in learning in the aggregate.

Other examples of counterintuitive results of the Model 1 approach can be easily generated. It may promote learning in the aggregate to reduce the course grade of students who refuse to purchase the required books in the course. I've been tempted to do this precisely because I have substantial anecdotal evidence that failure to acquire the required books in my courses is strongly correlated with lower grades. The threat of a substantial grade reduction for not having the required books would likely stimulate the desired behavior. However, once again, the student who learns well in spite of not having her own books is not treated fairly when her course grade is reduced.<sup>10</sup>

Finally, the instrumental perspective of Model 1 is without doubt deeply embedded in many common practices such as social promotion in grammar schools, grade inflation, punitive administrative grade changing, "E" for effort grading practices, and certain types of student development approaches on undergraduate residential campuses. However, such practices are not necessarily connected with the promotion of learning, *per se*. The defender of Model 1 can argue that the only commitment implied by Model 1 is a preference for learning over fairness, whenever the two are in conflict. That is, as long as a grading practice promotes learning, then learning trumps fairness.

What does the learning-trumps-fairness thesis entail? As with other applications of utilitarian reasoning, whether a specific reward/punishment grading practice actually improves student learning is an empirical issue. And, much of our pedagogy is based on our own experience—that is the nature of any professional art. However, when issues of fairness arise, it is certainly appropriate to require a higher standard of evidence than personal anecdotes.

Despite what my own teaching experience tells me about students who don't have the required books, the opportunities for bad inductive reasoning are lurking everywhere here. Correlation doesn't prove causation. Maybe students who don't buy books and subsequently get low grades can't read very well—the books are causally superfluous and owning them is a waste of money. Given the fairness stakes, proponents of Model 1 must be willing to swim in these deep empirical waters. If learning ought to trump fairness in grading, I would want experimental data that were established in valid and reliable ways. I would want the data to show an extraordinary benefit following from the unfair practice, not merely a statistically significant benefit. I would want a causal model that rules out other variables that could explain the correlation. And, I would want the benefit to be universal. If a single student were graded unfairly and received no learning benefit from the unfair grade, even though the rest of the students

---

<sup>10</sup>Although the reward-and-punishment grading systems associated with Model 1 are consequentialist by definition, *viz.*, the achievement of learning, nonconsequentialist justifications of rewards and punishments are common in ethical theory. For example, an instructor might increase the grade of a student who exhibited an unusual level of effort in learning the course material, perhaps working two jobs, and raising a family as a single parent. Such a reward might be made not as a means of increasing a particular type of behavior believed to be conducive to learning, but rather because such a student supposedly deserves the reward in some noninstrumental sense of "desert," e.g., a recognition of effort. Such a student might be thought by the instructor to deserve such recognition, pure and simple.

benefited, a red flag should be raised.<sup>11</sup> Most instructors—consequentialists and nonconsequentialists alike--will not want to throw even one of their students to the unfairness wolves in order to achieve a learning benefit in the aggregate. Punishing the innocent for the greater good is not a core moral principle in the academy. In short, learning doesn't always trump fairness and may regularly fail to trump fairness, once the empirical dust has settled.<sup>12</sup>

### **Model 2: Grades are the goal in the classroom.**

On Model 2, learning should be simply the means to accomplish the thing of true value, the grade, i.e., “good” grades are the purpose of getting a college education. This view receives reinforcement among students who seek scarce economic benefits that are distributed at least partly on the basis of grades. Highly selective colleges, for example, have traditionally placed substantial importance on high school grades as a criterion of admission and merit-based financial aid. So, good grades are the currency with which one can buy scarce resources and qualify for discounts on the price of those resources. The same can be said of the admissions processes of many graduate and professional programs. The fact that grades are at best merely a sufficient criterion of knowledge is easily lost on students who know that the proper badge is what will gain them access to scarce and desirable resources. It should come as no surprise that for the grade-grubbing student, what constitutes a “fair” grade is an A, and nothing less. Unhappily for us in the academy, this model finds a great partner in the students-are-customers model of higher education. Department chairs and deans who carefully review the course grade distributions of their faculty each semester no doubt reflect the pervasive influence of Model 2 in higher education.

### **The Common Error in Models 1 and 2.**

Academic grading does not differ in kind from a medical diagnosis, or from the grading of meats and produce. If I am evaluated by my physician and am told that I have shingles, am I *punished* by the diagnosis (a low “grade” of disease, rather than health)? I will surely be glum after my consultation, but the unhappiness I experience on learning my diagnosis is not the *purpose* of the diagnosis. For example, it would be preposterous for a physician to diagnose me with lung cancer when my lungs are healthy, simply to motivate me to quit smoking cigars. Constructing a lying diagnosis to increase patient healthiness is not a core moral principle in medicine. Is the grading of a particular steak as “prime” (the highest quality mark in the grading of beef) a reward, *per se*? The farmer may be quite pleased at the higher price fetched by the “prime” grade, but that happiness is not why meat inspectors grade meat.<sup>13</sup> Imagine the meat inspector who says that the most important goal of her job is increasing the quality and safety of meat. This inspector argues that by unfairly grading some producers' high quality meat as

---

<sup>11</sup> This might be viewed as a variation on Rawls' difference principle (1971, 60ff.).

<sup>12</sup> I am indebted to the editor for pressing the appearance of a conflict between the instructional goal of learning and grading fairly.

<sup>13</sup> Note the difference here between rewards/punishments and the concept of an award. Low grades are awarded just as are high grades, but presumably only the latter are rewarding.

unsanitarily and unhealthily produced, other meat producers are motivated to improve the standards of their operations. I argue that few of us would endorse such a plan.

So, Model 1, though subscribed to by many students and not a few faculty and administrators, simply confuses the psychological and social *consequences* of grading with the actual purpose of grading, i.e., that of providing the student and other transcript readers with an expert assessment of the student's knowledge or skills in the course subject matter.

The principle error in Model 2—"good" grades are the ultimate goal of education-- can also be understood with the medical diagnosis analogy. In one sense, my physician's diagnosis of shingles is certainly a consequence of my having shingles, just as my evaluation of the academic excellence of a student is a consequence of the student having learned the course material in a superior way. But, I do not acquire shingles in order to achieve the diagnosis, nor do I strive for excellent health as a means to getting a clean bill of health from my physician. As with Model 1, Model 2 confuses the *consequences* of grading with the *purpose* of grading. This is a very old distinction in philosophy, of course.

### **Model 3: Grading is an information process concerning mastery of course content.**

I argue that this is the most plausible model of grading. Model 3 focuses exclusively on the informational nature of grades. Grades recorded on official transcripts are read by many parties beyond the student, for example, faculty advisors, reference letter writers, employers, and graduate admissions committees. All transcript readers obtain information about a student's level of knowledge and competence in the academic areas under consideration by looking at the student's grades. That is, the kind of information that a grade communicates to the transcript reader is critical. Note that this is a normative claim, not an empirical one. The fact that a graduate admissions committee or a prospective employer may pay little attention to a student's transcript is not a counterargument to Model 3. If everyone suddenly stopped reading grade records and making decisions on the basis of grade records, this would not show that Model 3 should be rejected in favor of Model 1 or 2. Rather, it would show that the institution of grading itself had collapsed, thus mooted the entire question of fair and unfair grades. Of course, that is not the world we live in.

Now, contaminating the grade with information beyond the academic content of the course makes the transcript unreliable, even useless, in determining levels of knowledge and competence. Imagine a USDA meat grader who grades the farmer's beef carcasses in part on the great effort that the farmer expended in getting the beef to market or because of how much progress the farmer had made in raising beef cattle after a long career as a philosophy professor. Certainly, grading such beef "Prime" when it would otherwise not have been graded so highly is unfair. In this case, the grade of "Prime" is best regarded as an outright lie, and consequently a grave violation of professional duty (cf. Chartier 2003).

Assuming that Model 3 is the most plausible of the candidate models of grading, at least one principle of fair grading must therefore involve protecting the grade from contamination by irrelevant information. While we obviously not design our teaching solely on the goal of decreasing unfairness in grading, we should design our grading practices to limit the scope of what the grade reflects, viz., knowledge of course content and skills. To reject such a principle would entail a rejection of Model 3 in favor of Model 1 or Model 2. This point seems to resolve the conflict between the instructional goal of learning and the practice of grading fairly. While learning is—or should be--the instructional goal of every classroom, grading fairly is not an instructional goal, per se. At my institution, grading is a contractual obligation, and grading fairly is certainly an ethical obligation in the profession of teaching.

Finally, the three models are normative, not purely descriptive. Thus, to the critic who asks why we couldn't construct a model of grading in which grades are both informational *and* rewards and punishments, the answer is that we certainly can. In fact, that is likely to be descriptive of the way grades actually do function in our society. However, the concern here is not the empirical issue of how grades actually do function, but rather how they should function.

#### IV. Some Principles of Fair Grading

We know intuitively that grading our students is not morally neutral conduct, but what principles underlie our intuitions? The principles I will propose generally fall into two familiar groups: formal, and substantive. I borrow this distinction from justice theory in only a rough way to aid the discussion below. Principle 1 below, that grading should be impartial and consistent, is a formal principle that does not speak to the grading criteria themselves.<sup>14</sup> Likewise, Principles 1.1-1.6 below are corollaries of Principle 1 in that they speak to the distributive mechanics of grading. I do not argue that they are logical corollaries in a literal sense, although I think that they are derivative of Principle 1 in some looser way. Principle 2, that grading should pertain only to the student's knowledge of course content, is an elaboration of Model 3 above.

##### **Principle 1: Grading should be impartial and consistent.**

The intended scope of Principle 1 and the other principles that follow is limited to the individual instructor(s) in a given course. Even so, the relationship between grading and impartiality is not as clear as one might expect. Kenneth Howe states that in grading, "impartiality is the overriding moral principle" (1988, 325) and goes to say that "insuring impartiality is simply insuring validity: students ought to be evaluated in terms of publicly stated standards, not in terms of their looks, how nice they are, or how likely they are to give the instructor a bad time." I agree with Howe's statement here, but I do not agree with the implication that impartiality entails or excludes any substantive grading criterion. One may impartially administer preposterous grading criteria, just as one may administer the most sensible grading criteria in a prejudiced and unfair way. So what we should extract from Howe's claim at this point is just that impartiality is a central moral principle in grading, if not the overriding one.

Gregory Weis also seems to agree with the central role of impartiality when he observes, "I have a problem if I have no clear idea what grading is or ought to be, if I apply one theory of what grading is with this student, another with that student" (1995, 12). This statement appears to be an endorsement of Principle 1. But, contrary to Howe's claim that impartiality is crucial in grading, Weis then goes on to say that he is "leaving open the possibility that professors may take different factors into account when grading students" (ibid.). I think that what Weis means here is that different instructors may legitimately use different criteria in grading, not that one instructor may employ different criteria when grading different students in the same course. This is consistent with the scope of impartiality in Principle 1, viz., the individual instructor in a given course. If impartial and consistent grading is strictly course-relative, the same student would not necessarily be graded in a consistent manner across instructors. Principle 1

---

<sup>14</sup> Kenneth Howe (1988, 325) thinks that impartiality excludes certain substantive grading criteria such as appearance and friendliness.

could be made stronger so that inter-instructor, interdepartmental, and inter-institutional comparisons of grading are considered, but that would be a much more difficult task.

So, Principle 1 means that, all other things being equal, *nothing* should be relevant to any one student's grade in a given course that is not relevant to every other student's grade in that course. For instance, if "make-up" work is made available to one student who is ill or has a family emergency, it must be made available to all students under comparable circumstances. At many colleges, athletes receive special treatment with respect to making up course components that are missed because of away games, make-up games, play-off games, etc. Principle 1 says that comparable consideration must be extended to students who miss course components because of job travel schedules, sick children, etc. The fact that athletic absence is a consequence of an institution-funded student activity while taking one's child to the doctor is not, would not appear to be *prima facie* grounds for distinguishing between the two sorts of absences in terms of course work make-up opportunities.

Majors sometimes receive considerations not available to nonmajors. This practice may be justified in some cases, but only where the consideration has no impact on the course grade. For example, a chemistry major might be eligible to work as a lab assistant, setting up equipment, cleaning glassware, etc. This activity should in no way be factored into the student's course grade because it violates the general principle of impartiality. If it were to figure in the student's course grade, working as a lab assistant would be a grade component, but one which is not available to all students. Such a practice would be analogous to including a final exam in the Symbolic Logic course grade, but allowing only philosophy majors to take the final.

**Principle 1.1: Grade components should have determinate weights expressible as fractions of the final grade.**

Many professors mark student work on a raw point basis. For example, a course might have a total of 500 points, with two 100-point hour exams, a 100-point term paper, and a 200-point final exam. The weights of these course components are 20% for the two hour-exams and the term paper, and 40% for the final exam. If the professor takes class participation into consideration in determining the course grade, then by Principle 1.1, participation should have a determinate grade value, for example, 50 raw points. The 50 points can be extracted from the other grade components or added on so that the course would then have a total of 550 raw points. Either way, the respective weights will change. In the latter case, the correct weights of the grade components would be as follows: 18.18% each for the hour-exams and term paper, 36.36% for the final exam, and 9.1% for participation.

Another way of stating Principle 1.1 is that *whatever* a professor takes into consideration in assigning a grade should have a determinate grade value. Nothing in this formal principle *per se* would prohibit a professor from assigning a grade on the basis of the students' height, but only that the criterion must have a determinate value in the students' composite course grade. If I simply tell my students that I will take class participation into account in the course grade, but I don't tell them what the weight of that component is, I have not satisfied the requirement of Principle 1.1. If I weigh participation differently for different students, then I have violated Principle 1 in its most general form: I have failed to be impartial.

**Principle 1.2: Grade components and their weights should be published at the beginning of the term.**

---

This corollary to Principle 1 is intended to exclude all types of *ex post facto* grading policies. For example, the student who participates frequently should not receive a higher course grade than another student with the same marks *unless* class participation is an explicit grade component of which everyone is aware at the beginning of the term, and for which there is a determinate grade weight (Principle 1.1).

Principle 1.2 is a rule that I sometimes find necessary to bend a bit, viz., with mid-semester revisions of course design, while still observing the general requirement of impartiality. For example, I might reduce the number of short papers from five to four because of days lost to unforeseen events, bad weather, an unplanned work assignment that takes me off-campus, etc. If the five papers were collectively worth 40% of the course grade, this means that each of the papers already completed will increase in value from 8% each to 10% each. Because such a change may place some students at a grading disadvantage relative to others, it is a potential violation of both Principle 1 and 1.2. In this case, I discuss the possible change with my students and ask them to review the change prior to a vote at the next class meeting. In order for the change in course design to occur, the students must support it unanimously. If a single student present and voting believes that his or her course grade will be harmed by the change, nothing happens. The five papers will be required as originally scheduled.

Principle 1.2 does not prohibit using “pop” quizzes or other unannounced graded activities in a given lecture. Some instructors may find that students are more motivated to prepare for lecture if there is the prospect of a pop quiz, for example. This principle merely requires that the instructor publish the weight of the pop quizzes at the beginning of the semester. Note that the justification for publishing the pop quiz weights at the beginning of the semester is not to increase student preparation for lecture, though it may have that effect. Instead, as with all of the other course grade components, advance publication of grade components contributes to a meeting of the minds between professor and students about the precise conditions by which the course grade will be constructed.<sup>15</sup>

**Principle 1.3: “Forced” grouping of a set of scores into As, Bs, Cs, Ds, and Fs, is an inherently unfair method of grading.**

What I mean here is the assumption that *every* set of scores contains some Fs, some Ds, some Cs, some Bs, and some As, and that grading consists basically of locating the dividing lines. Such a methodology is grounded in the concept of relative peer ranking and its most common form is sometimes called “grading on the (normal) curve.” One reason that grading on the curve is unfair is that grade distributions cannot be guaranteed to be normal even when the sample size is large. Upper division courses, for example, can easily consist of As and Bs only, or even entirely of As. Skill-based courses such as mathematics, logic, computer programming, and foreign languages can display what is known in statistics as a bimodal pattern of distribution, i.e., most of the students are clustered at the top and bottom of the grade range, with only a small proportion of Cs. Grading on the normal curve in the proper sense of the term may be uncommon in undergraduate grading, but many faculty use some type of peer ranking mechanism in determining course grades.<sup>16</sup>

---

<sup>15</sup> Some instructors like to involve their students in the determination of course grade components. This presents no inherent issues of fairness beyond the principles presented here. Put simply, neither the instructor nor the students should construct unfair grading practices.

<sup>16</sup> An anonymous reviewer observes that college catalogs sometimes define letter grades in terms that are clearly relative rank-oriented, e.g., a C grade that is defined in terms of “average work.” This is

---

Another reason that peer-ranked grading is unfair is that such practices can deprive students of official recognition of high academic achievement, i.e., a high level of mastery of course content. If a student's evaluation places her at the bottom of the curve, then she will fail the course, regardless of the degree to which she has mastered the course material. If one objects, "Well, that never happens in my courses because the students who get Fs deserved them," we know that there must be some *implicit* system of evaluation in operation beyond relative peer ranking. The objector evidently has a valid grading instrument by which the results of grading on the curve can be confirmed or disconfirmed. The solution is to entirely eliminate the peer ranking system in exchange for the valid instrument operating in the background.

Finally, violating Principle 1.3 means that the course grade will explicitly fail to communicate the degree of mastery of course content. The idea that we *know*, prior to any evaluation of student work, that some students will receive Fs in the course, some students will receive As in the course, etc. is profoundly unempirical. Worse, it is contrary to the idea that above all, grades reflect the student's knowledge of course content (see Principle 2 below).

**Principle 1.4: If temporal sequencing of a given student's grades can have an impact on the course grade, that impact should be built into the weights of the grade components and applied to all students.**

Suppose a course grade consists of scores on five exams of equal weight (20% each), evenly spaced throughout the course. If the instructor believes that Smith, who has scored B, B, B, A, A, in that order, should receive an A in the course, but Jones, who has scored A, A, B, B, B, in that order, should receive a B, then the instructor is implicitly assigning greater weight to exams taken later in the course. Principle 1.4 would require those weights to be stated explicitly, e.g., 10%, 15%, 20%, 25%, 30%. This principle would also apply to the practice of "giving the benefit of the doubt" to the student who finishes the course with a high B average (including the final exam), but, by virtue of having written a solid A on the final, receives an A for the course.

Complying with this principle is most troublesome in sequential courses where cumulative knowledge and skills are central to the mastery of course content, e.g., conversational foreign language, mathematics, and logic courses. In such courses, what the student knows at the end of the course is really what counts in evaluating mastery of course content. In philosophy, logic courses will usually have this feature, but so may other courses where the mastery of several competing theories is best revealed at the end of the course, and not in piecemeal fashion throughout the semester. If the instructor believes that how quickly the student arrived at her final level of competence, how many false starts she had along the way, etc., are of secondary importance, the solution is to place great weight on the final exam and relatively little weight on the quizzes or hour exams during the semester.

---

certainly true and seems to be a reasonable way to define academic grades in the large—even if "normal" is so vague as to be unusable in practice. But Principle 1.3 is concerned only with relatively small sets of scores generated by a grade component in a given course—it is not intended to be a counterargument to the law of large numbers or the central limit theorem. Just as the meat grader cannot presume that a visit to a given packing plant will generate meat grades across the entire grade spectrum, an instructor cannot presume that a given sample—an hour exam, say—is sure to contain academic grades across the entire grading spectrum.

**Principle 1.5: Every student must receive a mark for every grade component in the course.**

Principle 1.5 means that under normal circumstances, no student may be exempted from being evaluated with respect to a given grade component. To do so would be to violate the principle of impartiality. There are several types of violation of this principle in academe, but three will be familiar to most instructors: “adequate information” grading exemptions, “motivational” grading exemptions, and grade dropping. The practice of extra credit may also be reviewed under Principle 1.5. Although it does not intrinsically involve violating this rule, extra credit can easily veer into foul territory.

**Grade Exemptions**

There are three sorts of widely practiced grading exemptions. First, there is the practice of exempting “A” students from taking the final examination or completing some other grade component because the instructor believes that the evaluation of that student’s progress in the course is complete. In this practice, there is no need for further evidence of the A student’s mastery of the material. This belief may well have strong justification. In fact, in some of my courses, the student’s midterm grade is such a strong predictor of the course grade that I could easily never grade the final exam and still assign highly accurate course grades.

However, there are problems with this statistical approach to fairness. One problem is that there are no perfect predictors of the future. So, even though an A student is likely to prepare the material for the final exam to an A level, there is no guarantee that he or she will. Late semester “melt-downs” do occur. In such a case, the instructor will have based that student’s course grade on inadequate information.

Here’s another problem. What we should do if the exempted student chooses to take the final exam anyway and does so poorly that the score lowers her course grade to a B? Should the instructor simply ignore the final exam? To do so would contaminate the grading process, unless the instructor ignores the final exam for every student where the final exam score would reduce the course grade below what the student’s course average was going into the final. Either the final exam is essential to determining the course grade, or it is not. There is no in-between. If we say, “Sometimes the final exam is important and sometimes it isn’t,” what we probably mean is that impartiality is getting in the way of our instinctive judgments about who should receive what course grade. Such instincts are an impediment to fairness and their influence should be eliminated wherever possible.

One might argue that the exemption from the final exam for A students is available to everyone and is therefore fair. But if an A student can be exempted from the final because the information gained by the instructor will be superfluous, then perhaps the same is true of B students, and perhaps C students as well. Indeed, one can imagine courses in which the student’s course average prior to the final exam is highly predictive of the course grade for all grade categories. So how can it be fair if the policy is extended only to A students? The answer is, it’s not fair at all. A fair exemption policy would extend the exemption to all grade categories, not just some grade categories. For example, if Donna has a C average going into the final exam, then she may skip the final and automatically receive a C for the course. That is, regardless of your course average prior to the final exam, if you do not take the final exam, then your course grade will be whatever course average you had prior to the final. However, once the exemption policy has been made equitable in this way, the final exam is now clearly superfluous in determining the

course grade, and so should be dropped completely from the course design. Once again, either the final exam is essential to determining the course grade, or it is not. Pick one, and apply it uniformly.

A second type of exemption from course components is where the instructor believes that the exemption is a prize that will motivate students to excel. This is an empirical claim that may or may not be true. Let's suppose that an exemption from the final for students going into the final with an A average does in fact demonstrably motivate B students to become A students. Does this validate the practice? The answer lies in whether we should construct grading practices on instrumental grounds. I will argue in Principle 2.2 below that we should not.

The third common sort of grade exemption is to drop the lowest exam, quiz, or paper grade from the calculation of the course grade. For example, a course might have four quizzes, each worth 10% of the course grade, and a final exam worth 60%. Under the drop practice, the instructor eliminates the quiz with the lowest mark from each student's course grade. Typically, different students will have different quizzes eliminated, but that is not an essential feature of the practice. The important point here is that the practice basically takes the three remaining quizzes—and possibly other grade components—and *reweighs* them.

Reweighting can be done by several methods. First, the instructor can reweigh the three remaining quizzes to have the same collective value as the original four quizzes, i.e., 40%, so that each quiz acquires a new individual weight of 13.33%.

Second, the instructor may simply remove the dropped quiz value from the denominator of the course average, so that, in effect, there are now 90 standard points possible in the course instead of 100. This method reweighs the three remaining quizzes to a new individual weight of 11.11%. And, since the course average denominator has been reduced from 100 to 90, all other grade components are automatically reweighed. In our example, the final exam would become valued at 60 standard points out of 90, or 66.67%, instead of the original 60%.

A third possibility is that the instructor does not reweigh the remaining three quizzes but instead adds the dropped quiz weight to some other grade component, e.g., the final exam. In the example, this would increase the final exam weight from 60% to 70%.

There is a fourth possibility, although we can hope that it is avoided. In this case, the instructor drops the lowest quiz score *and* leaves the denominator of the course average unchanged at 100% (in standard points) *and* does not move the dropped weight to another grade component. Since this has the effect of penalizing all students by the amount of their dropped score—thus “curving” all course averages down—I will assume that all instructors who use the dropped score grading mechanism employ one of the first three reweighing mechanisms.

Why is grade dropping unfair? After all, students rarely complain about it. The thoughtful student might understand that the reduction in the number of grade components and the consequent increase in the weights of the remaining components just might not be to his or her advantage. However, it's more likely that the typical student is completely unaware of the implicit reweighing of other grade components. This type of exemption can thus be seen as a likely violation of Principle 1.2 above, viz., that grade components and their weights must be published at the beginning of the term. Still, this requirement can be easily met. If so, then what?

The core problem with the ordinary sort of grade dropping is that students are actually not being treated equitably. When an instructor designs grades components such as quizzes, the assumption is that he or she believes those components are each essential to the determination of a course grade. For example, in an Ancient and Medieval Philosophy course, I might have four quizzes over the pre-Socratics,

---

Plato, Aristotle, and Epicurus, respectively. One of the goals of the course is that students learn material in each of those four areas.

Grade dropping, by definition, involves a typically random exclusion of any one of several course components, each of which are purportedly essential to the course grade. By “random,” I mean that the probability that any one of the eligible grade components will in fact be excluded is roughly the same among the eligible components. The exception to this is when one of the four quizzes, say, is either so poorly written, so difficult, or for which the students are so poorly prepared that it is the lowest-score quiz for most students. In the usual case, where the four quizzes are roughly equivalent in composition, difficulty, and student preparation, the lowest-score characteristic will be held in roughly equal numbers by each of the four quizzes.

So, in the example above, none of the four quizzes is protected from exemption as a course grade component. Expressed as an explicit contradiction, Quiz 1 (pre-Socratics) may be essential to Student A’s course grade, but not essential to Student B’s course grade. Quiz 2 (Plato) may be essential to Student B’s course grade, but not essential to Student A’s course grade, and so on. So, is Quiz 1 essential to the determination of the course grade or not? What about Quiz 2? If one of the goals of the course is that my students learn content in each of the four areas, I am arguing that it is *prima facie* unfair to say to one student, “You know little or nothing about Plato, but that fact won’t enter into your course grade,” while I say to another student who has received *precisely the same mark* on the Plato quiz, “You know little or nothing about Plato, and I’m going to include that fact in your course grade.”<sup>17</sup>

The only difference between the usual sort of grade dropping and the deliberate application of different grading criteria to different students is that grade dropping is typically a randomized selection of the differing grade components, rather than one that is based on, say, athletic participation or physical attractiveness. Despite the apparent equitability of randomness, grade dropping allows one student to eliminate an unflattering but accurate assessment of his or her knowledge from the course grade while another student is required to accept perhaps precisely the same assessment as part of the course grade. The fact that all students are wronged “equally,” does not justify the practice.

## Extra Credit

How does “extra credit” work stack up against Principle 1.5? The principle requires that every student must be given the opportunity to complete work for “extra credit.” It must be announced at the beginning of the semester and, since it is a grade component, it must be given a specific weight in the course grade. This prevents the very real case of the star athlete failing your course, who now *must*—says the department chair, or the dean, or the provost—be given some “extra credit” work. “Yes,” you can then say. “All of my students can submit work for ‘extra credit’.” The Dean is happy and you have not violated the principle of impartiality. But, there is more to be said about “extra credit” and “make-up work.”

I use scare quotes around “extra credit” because the term itself is likely a misnomer. Arithmetically, there simply is no such thing as extra credit, although one species of extra credit is

---

<sup>17</sup> Grade dropping does not seem to be *prima facie* unfair in the following case: there are three quizzes on Plato, all roughly identical to each other. The highest score of the three quizzes will be entered into the course grade and the other two scores will not be included. That is, the student has three chances to demonstrate knowledge of the required content.

certainly a form of grade inflation. Assuming compliance with Principle 1.5, “extra credit” is just another grade component that students may choose to complete. In its most common form, what so-called extra credit does is to simply recenter the grading scale. For example, suppose that a course has 100 raw points. The instructor adds an “extra credit” paper worth 10 raw points. This simply means that there are now 110 raw points possible in the course, so student who receives a perfect score of 110/110 has a 100% course average. The student who achieves 99/110 points has a 90% course average, the student who achieves 88/110 points has an 80% course average, etc.

Some instructors will object, “Well, that’s not what I mean by ‘extra credit’. What I mean is that the student with a perfect score of 110 raw points will have a 110/100 or 110% average.” But, this is just a round-about way of saying the same thing arithmetically, *assuming* that the instructor adjusts his or her grading scale accordingly.

Suppose my objector says, “No, you still don’t have it right, because I leave the denominator at 100 and add the extra credit points only to the numerator, *and* I don’t adjust my grading scale accordingly, and this is what ‘extra credit’ really means.” Now, I have no evidence of the extent of this sort of extra credit, but it is clearly nothing more than what is usually meant by “curving,” yet another form of recentering grades. Curving usually involves adding points to a course average in order to force the class mean to rise, e.g., to the midpoint of the instructor’s “C” range. (Curving of this sort is not to be confused with grading on the curve, as discussed above.)

There is certainly nothing inequitable about any of these variations on extra credit as long as they conform to Principle 1.5, but the last species discussed is just a convoluted form of grade inflation and as such, certainly deserves review.

**Principle 1.6: Grade components should be designed to be performable by a wide variety of students.**

This principle is one that must be tailored to the specific student population in the classroom and is likely to be followed faithfully by virtually all instructors. But, as best as we may try, sometimes our grade components are not accessible to some students unless we make accommodations. This simply means that differential treatment is sometimes compatible with the principle of impartiality.

For example, a visually impaired student may need extra time to write an examination. The student might not be able to write the exam by hand and instead must be provided with a computer. With the recent growth in diagnoses of learning disabilities, I can expect to have at least one student who will require some sort of accommodation, perhaps nothing more than a little extra time. I generally allow foreign students to use an English dictionary unless I’m absolutely certain that it would provide an unfair advantage to the student over the native English speakers. Principle 1.6 may also be the grounds for accommodating working students, student athletes, and other types of “handicaps” that require the instructor to make adjustments in the normal methods of assessment. As A. D. Woozley points out, “[w]hat is unjust is the denial of an asked-for opportunity to display an ability, if there is some evidence of its possession” (Woozley 1973, 117).

Let us say that in every case of a *fair* accommodation, the result is to bring the student into a position of rough equality with the other students. An *unfair* accommodation is then one which provides an overall advantage to the student over the other students rather than equalizing the students relative to each other. Some excused-absence policies are unfair accommodations. For example, institutional policies that excuse certain types of absences but not others are not only cumbersome and labor-intensive, but fundamentally flawed by the assumption that some absences should not to be excused. If an absence

---

policy punishes the working parent with a sick child or an unavoidable work obligation, and excuses the basketball player, the choir member, the college theatre usher, the student on a field trip in another course, etc., the policy is unfair. Such policies are certainly not impartial. The reason for an absence is generally irrelevant to any education consequence of the absence.

There are other types of accommodation that clearly cross the line of acceptability. Any kind of accommodation that essentially exempts the student from completing a given grade component, thus violating Principle 1.5 above, is suspect. For example, the baseball player who misses the written final exam because of a rescheduled rain game should be given an opportunity to take the final exam at another time. But, that student should not be allowed to substitute a 15-minute oral exam in its place unless every other student is also given the opportunity of substituting an oral exam for the written exam.<sup>18</sup>

On the other hand, a student who is hospitalized for three weeks but is able to keep up with her homework should be provided opportunities to be evaluated even if those methods necessarily must depart from the usual grade components. For example, I do not normally allow students to make up missed group activities, such as debates. If you're absent for any reason, that's too bad. I may be able to reschedule you with another group, but barring that kind of accommodation, you must participate in a debate in order to receive a debate grade. The hospitalized student situation challenges Principle 1.5 directly and may require an extraordinary accommodation, for example, a one-on-one debate with the instructor. The key point is that this is an extraordinary accommodation, not an ordinary one.

To conclude the discussion of the general principle of impartiality, we have identified certain common grading practices as unfair, viz., grade component exemption, treating some students with partiality in absences, make-up opportunities, etc., failing a student solely because he or she is at the bottom of the class grade distribution, using "secret" grading criteria such as participation that have no determinate, public role in the course grade, and making accommodations for students where none are required or where the accommodations give those students an unfair advantage over other students who do not receive the accommodation.

**Principle 2: Grading should be based on the student's competence in the academic content of the course.**

This is the key substantive condition of fair grading (necessary but not sufficient) and is, without doubt, the most complex from an instructional perspective. However intuitive this principle may seem, I will argue that many common grading practices are ruled out if one adheres to the principle.

Versions of Principle 2 can be found in the literature on grading. For example, John Sabini's and John Monterosso's scenario study of student opinions about fair grading refers to this principle as the "strict performance model" of grading, describing it as "the idea that grades should be determined by performance alone . . . [and giving] no place to moral worth in assigning grades" (2003, 191).

In her classic book, *Tools for Teaching*, Barbara Gross Davis lists as a general grading strategy the injunction to "[g]rade on the basis of students' mastery of knowledge and skills." She goes on to say that the use of nonacademic factors such as "classroom behavior, effort, classroom participation, attendance,

---

<sup>18</sup> Consider the widely reported story of Ohio State University's star football player, Maurice Clarett. Clarett was given special oral exams in a course where he walked out of the regular midterm exam and did not take the regular final exam. Clarett is reported to have been the only student among 80 in the class who was given oral examinations (Freeman 2003).

punctuality, attitude, personality traits, or student interest in the course material . . . obscure the primary meaning of the grade” (Davis 1993).

Gary Chartier calls Principle 2 the “principle of academic exclusivity.” Chartier’s version “requires that, as far as possible, all nonacademic factors be excluded from consideration when instructors determine grades” (Chartier 2003, 39). Chartier’s version of Principle 2 appears to be motivated in part by legal rulings such as a finding by a Federal District court that a “rule that calls for a grade reduction to discipline nonacademic conduct is illegal, and null and void” (*Smith v. Sch. City of Hobart* 1993, as quoted in Chartier 2003, 38). I do not offer the law as proof of a moral principle here. What is instructive is that to grade fairly is to meet a moral obligation to the student. Failing that obligation is a violation of the correlative right of the student to be graded fairly; a right that students do in fact pursue in law. Forewarned is forearmed.

Principle 2 also captures the heart of Section II of the AAUP’s *Joint Statement on Rights and Freedoms of Students*<sup>19</sup> (American Association of University Professors 2001) that forbids “prejudiced or capricious academic evaluation,” and requires that “[s]tudent performance should be evaluated solely on an academic basis, not on opinions or conduct in matters unrelated to academic standards.” Here, too, it is not surprising that such statements have shown up in legal opinions. For example, the *Statement on Professional Ethics* asserts in part that “[p]rofessors make every reasonable effort . . . to ensure that their evaluations of students reflect each student’s true merit” (AAUP 2001, 133). This statement was found by the 7th Circuit (*Keen v. Penson* 1992) to be operative in the findings of a faculty committee where the AAUP statement had been made part of the university rules governing faculty conduct (AAUP 2001, 309).

Since course objectives could include nonacademic factors such as attendance and comportment, Principle 2 simply asserts that the course grade must be restricted to the student’s competence in the academic objectives. Another example would be a faculty policy that the institutional mission goals be reflected in the course objectives for every course. Suppose that “respect for diverse religious beliefs” is one of the institution’s mission goals. The syllabus for Cell and Molecular Biology might include this goal as a course objective, as a matter of faculty policy, just as every other course in the catalog does. Principle 2 simply means that the course grade for Cell and Molecular Biology will reflect only the student’s competence in the academic content of the course, viz., cell and molecular biology, and will not reflect other nonacademic institutional, departmental, or instructor goals.

The point here is that instructors can build all sorts of institutional objectives into their courses that are subsequently assessed in some way without those objectives being reflected in the course grade. This shows that Principle 2 is compatible with *ungraded assessment* of noncontent-related objectives in the classroom.

Finally, Principle 2 clearly excludes standard sorts of discrimination against members of protected classes under Federal and state law, as well as other forms of discrimination not legally recognized, e.g., weight and physical appearance. I will assume here that these forms of unfairness in grading are uncontroversial and will not discuss them further.

Several corollary principles associated with Principle 2 follow.

---

<sup>19</sup> Endorsed by the American Association of University Professors, United States National Student Association, Association of American Colleges and Universities, National Association of Student Personnel Administrators, and the National Association for Women in Education, among others (AAUP 2001, 261-267).

---

**Principle 2.1: Grades should be assigned on the basis of an expert evaluation of student work.**

To assign a grade to a student without examining student work is a violation sine qua non of course-relevant grading. The old joke cited earlier about grading exams by throwing them down the stairs would be an easy example of a violation of 2.1. The silliness of the grading process arises out of the absence of any course-relevant criteria. The requirement of instructor expertise in the subject matter is necessary, but not sufficient, to satisfy Principle 2.1. I need to actually employ my expertise in the subject matter when I grade my students. Because there are several departures from Principle 2.1, let's examine them separately.

**Administrative Assignment of Grades**

The administrative assignment of grades without examination of student work by faculty in the relevant subject area is not common, but some of us have witnessed such actions. The AAUP statement, *The Assignment of Course Grades and Student Appeals* (American Association of University Professors 2001), is instructive here:

The faculty member offering the course . . . should be responsible for the evaluation of student course work and, under normal circumstances, is the sole judge of the grades received by the students in that course. . . . Under no circumstances should administrative officers on their own authority substitute their judgment for that of the faculty concerning the assignment of a grade.

**Peer Grading**

By peer grading, I mean the practice of requiring students to summatively evaluate<sup>20</sup> one another in a given course. In other words, the instructor uses the results of those peer evaluations as one of the components of the course grade. This sort of peer evaluation must be distinguished from formative peer evaluation in which students engage in the evaluation of each other's writing or other work in order to facilitate their own learning (for example, see Wilson 2006). While I grant that it is possible to construct a method of peer grading that meets at least some of the requirements of impartiality, it would require empirical research to verify that one's students were actually unbiased in their evaluations of their fellow students. At any rate, it still remains that peer evaluations, by definition, are not expert evaluations, at least with respect to the course content. This is a fatal flaw in the practice of peer grading.

On the other hand, the practice of using student "graders" doesn't appear to entail a violation of Principle 2.1. Many of us had our first experience with student graders in grammar school when we each passed our spelling quizzes to the student behind us and then graded our classmates' quizzes under the direction of the teacher. In the undergraduate setting, student graders are undergraduate students with an advanced knowledge of the material—typically in lower level courses—who apply a grading rubric

---

<sup>20</sup> Michael Scriven (1967) formulated the distinction between "formative" evaluation in which the goal of evaluation is developmental, that is, to help revise and improve an ongoing process such as teaching or learning, and "summative" evaluation which occurs after a process ends and is typically used by decision-makers in faculty retention, by faculty in assigning course grades to students, etc.

---

developed by the instructor. This practice imitates the role of teaching assistants in graduate programs, but normally does not call on the student grader to develop the grading rubric. The student grader performs a purely mechanical task of applying the professor's grading rubric to student work.

## **Comportment**

In the context of pre-college education, comportment concerns student behavior that does not conform to school rules. This includes poor conduct on the bus and playground, in the gym, hallways, and cafeteria, and in the classroom. Misconduct can include absence from school, tardiness, fighting, inappropriate language, defiant behavior, and dress/grooming code violations. Many primary and middle schools have a comportment (or deportment) grade that is separately reported on the grade card. A sufficient level of misconduct can result in detention or even expulsion.

Similarly, the typical undergraduate campus has a formal system of monitoring and punishing non-classroom conduct that violates institutional rules. Those systems are typically administered by Student Affairs and can involve a variety of sanctions, including expulsion. The most likely sorts of comportment that are monitored within the college classroom are attendance, tardiness, and private conversations, and many college faculty have conduct rules concerning wearing hats, using cell phones in class, using earphones in class, and other behaviors regarded as inappropriate. But regardless of the nature of the comportment rules themselves, the question here is whether comportment itself should be a course grade component.

While we may complain of the failure of some of our students to observe common courtesies in the classroom, and while some of us may even demand that students comport themselves in certain ways, e.g., "Remove your hat when you enter the room," I suspect that few of us would argue that students should be sanctioned—let alone graded—on such criteria as style of dress, hairstyle, or jewelry. Still, one could imagine some sort of content-related comportment grade in a management course in human resources. The professor might grade students on their demonstrated knowledge of the rules of etiquette in anticipation of their future treatment of employees in the workplace, e.g., "When you play the role of interviewer in the mock interviews, you must wear business attire and remove nose, tongue, and eyebrow piercings."

On the other hand, it is hard to imagine any justification for any sort of comportment grade in a symbolic logic course. This is not to say that faculty may not rightly exert any control over comportment. Indeed, the instructor has a professional responsibility to maintain an appropriate academic environment. If Ralph insists on arriving at lecture completely nude and thereby disrupting class, he may be expelled from the classroom (for disrupting class at least, if not for the lack of clothing). But his disruptive conduct cannot fairly be factored into a formal assessment of Ralph's understanding of symbolic logic. His conduct is unacceptable—possibly illegal—and cannot be tolerated because it interferes with the rightful access to instruction by the other students, but his conduct, per se, has nothing whatever to do with his understanding of the subject matter of the course.

## **Tardiness**

Tardiness is a species of comportment. A colleague who teaches at a large university in New York has pointed out to me that grading on tardiness proved to be an effective control over the disruption caused by late arrivals. In a lecture hall of 300 students, the doors at the rear of the hall would open and close loudly several times during the first few minutes of lecture as tardy students entered the hall. Each

time the doors made a noise, students' heads would turn towards the rear, disrupting lecture. After instituting a stiff grade penalty for tardiness, late arrivals virtually disappeared.

Regardless of its effectiveness, I argue that such a grading practice is unfair. The time of day that a student enters the classroom has no intrinsic connection with her understanding of the course material. This conclusion of unfairness should not alarm us, however, since there are many examples of social policy that may be very effective in achieving the desired end, but are nonetheless unfair or unjust, e.g., warrantless sidewalk strip searches for illegal drugs (yielding high arrest rates in certain parts of town), basing college admissions solely on high family income (income and academic success are strongly correlated), etc.

Specifically, there are methods other than grading by which we can control comportment that disrupts the academic environment. For example, admission to a lecture hall can be physically barred to tardy students, just as concert-goers may be physically barred from entry to the concert hall in the middle of a musical performance. Nonacademic sanctions can be used to deter tardiness such as placing the tardy student at the end of the line in next year's dormitory room lottery. Tardiness could be reported to coaches, persons who, in my experience, take a very dim view of that behavior in their athletes. Such mechanisms may be institutionally inconvenient, compared to reducing a student's grade for tardiness, but inconvenience is clearly not a *prima facie* justification for failing to do what is right. Unfortunately, if the institution refuses to support the instructional environment in order to deter tardiness, the professor is caught in the middle. If the professor has no alternative control of disruptive tardiness, she may be forced to use grading as a behavioral control on purely utilitarian grounds—it's the lesser of two evils. The point here is that such a grading criterion is unfair because it violates Principle 2.1. Whether or not its unfairness entails an absolute ban in every conceivable case is another matter.

## Attendance

Attendance is probably the most common form of comportment that finds its way into course grades. Grading on attendance can be reviewed under Principle 2.1 along the same lines of argument as grading on tardiness and other forms of comportment. By "grading on attendance" I mean grading policies that treat attendance, *per se*, as a component of the course grade. For example, an instructor might have a policy that reads, "Attendance will affect your grade. Three unexcused absences are allowed without grade penalty, but each absence beyond three will result in a reduction of one half of a letter grade in your course grade." Another sort of policy might read, "Attendance will comprise 10% of your course grade. Since there are 28 lectures in the course, attending a lecture contributes about 0.36% towards your course grade. All absences reduce your course average, regardless of the reason."

Principle 2.1 entails that such grading policies are unfair. The reason is simple: mere physical presence in the classroom does not constitute "student work." And, as the sample policies above indicate, such policies explicitly provide a grade component without the examination of any student work. This result will strike many teachers as counterintuitive simply because most of us feel that attendance in class makes an important contribution to student achievement. Why else would we have heated debates in faculty committees and departments about attendance policies and retention, verification of excused absences, grading on attendance, etc., if attendance didn't matter? However, attendance in class is neither necessary nor sufficient for learning (or if it were necessary, we would have a knock-down argument for denying regional accreditation to all distance education programs). Attendance *per se* is clearly not a valid instrument for measuring course content knowledge and competencies, and therefore, it cannot fairly play a role in the course grade.

---

This is not to say that attendance should never enter into our thinking about a student's academic achievement. Consider the case of academic dismissal from the institution. In a vast majority of student appeals of academic dismissal that I've heard over the years, a standard condition of returning to the college is regular attendance in class. At my institution, this is monitored by the director of our Academic Success Center, who reports compliance back to the faculty committee that hears dismissal appeals. Such a mechanism is entirely different than actually grading the student on attendance as if mere attendance was somehow a valid instrument by which knowledge and competence in course content could be measured.

By Principle 2.1, we should not support readmission conditions that would fail a student in a course or reduce her grade merely because she cut class beyond the prescribed limit. As in the case of tardiness, there are nonacademic sanctions that can be imposed on students for whom mandatory attendance is administratively imposed. Bar the student from athletics, extend her academic probation, require her to attend academic counseling, or even expel her from the College, but don't contaminate the instructor's evaluation of her level of knowledge in the course content with an attendance grade component.

Some might argue that there are certain courses where attendance is necessary in order to learn a specific skill. Musical performance, science labs, and stagecraft are obvious examples. But so are philosophy courses, as well as other subjects where competency in listening to the oral arguments of others is an explicit course objective. Since one must be present to play an instrument or sing, perform experiments, build a set, or listen to the philosophical arguments of others, my objector would say that attendance in such courses can therefore be justifiably calculated in the course grade. However, this objection fails. Attendance, per se, is only a necessary condition for performing those various skills--it is not sufficient. Consequently, what we need are other metrics that are sufficient to measure playing an instrument, experimenting in the laboratory, building a set, and critically listening to oral arguments. In such courses, the grade component should concern the performance of the skill itself, not simply being available to perform the skill. Did the student play the composition from memory? Did the student correctly perform the laboratory analysis of the unknown substance? These are questions regarding skill assessment that presuppose physical presence in the classroom but cannot be answered by merely observing that the student is physically present in the room.

Another objection that might be raised regarding grading on attendance is that if a student is not present then she is less likely to learn the material covered in the lecture. This is an empirical question, of course, though most instructors no doubt subscribe to the belief that learning actually takes place in the classroom. Nonetheless, the objection ignores the nature of fair grading. Grades presumably should not reflect what merely *might* be the case, even if the probability is high. A fair grade reflects an expert assessment of the student's actual achievement. Grades are not short-hand for what the student might know, probably knows, or ought to know, if she attended class.

Fortunately, removing attendance from one's grading is very easy. For example, if a course includes in-class exercises such as group writing, case presentations, commenting on another student's presentation, etc., each of these activities can be given a determinate weight in the course grade. Since physical presence in the classroom is necessary for the performance of in-class activities, the student who cuts class obviously receives a zero for that in-class grade component. This approach is quite different than grading on attendance, per se, where mere physical presence in the classroom is assigned a specific weight in the course grade, in addition to whatever in-class activities have been included in the grade. In place of mere attendance, we must instead develop an appropriate rubric for evaluating the in-class activity.

---

## Moral Virtues

What can we say about grading other conduct that is not intrinsically related to course content? We need only ask ourselves whether that conduct is “student work” in the meaning of Principle 2.1. For example, I have an explicit rule in my courses that we must all show each other respect in our discussions. However, I do not grade students on their compliance with this rule. Good philosophy can be done rudely. Human decency, or the lack of it, has no effect on the soundness of an argument. Even though good manners in philosophical argument is a moral value that I’d like my students to embrace, I’m not going to grade them on it, nor should I. Students who fail to comply with the rules of civilized discourse can be advised and counseled, or even barred from attending lecture, but to build personal attributes into a course grade requires a content-based rationale not available in more than a few courses.

More generally, we should be careful not to increase a student’s course grade on the basis of his or her cheerfulness, helpfulness, dedication, sensitivity, and other moral virtues, nor reduce a student’s grade for lack of those virtues. These are dimensions of student conduct that can be very seductive when it comes the assignment of a grade. After all, many colleges assert the moral development of the student as an explicit institutional objective. Nonetheless, whether such moral development is relevant to the content of a given academic course is largely an empirical matter. For example, moral development/sensitivity, e.g., compassion, may be specifically relevant to a major in the allied health professions, but it presumably has no relevance to course content in a history of philosophy course. Generally speaking, cheerfulness is typically not course content relevant, and it is patently unfair to grade a student on her cheerfulness in such courses.

Again, Principle 2.1 does not forbid pursuit of virtuous conduct in the classroom, but merely the inclusion of it in the course grade. So, this principle is indifferent to the perennial question in philosophy and other disciplines about moral advocacy versus moral neutrality on the part of the instructor. It should also be noted that this discussion is distinct from the ancient question, e.g., in the *Meno*, of whether or not virtue can be taught. My point is limited to the wrongness of grading students on their virtues when acquiring those virtuous characteristics is not relevant to the course content.<sup>21</sup>

### **Principle 2.2: Grades and grading practices generally should not be based on instrumental grounds.**

This corollary to Principle 2 is more controversial, but is clearly implied by it. If grading is a evaluation of the student’s competence in the subject matter of the course (“learning objectives” in educational assessment jargon), then grading cannot normally be a means to some other end. By “end” I mean “goal,” e.g., to encourage or motivate students, or to reward or punish. Neither the individual instructor, a department, or the institution should use grades to pursue other goals no matter how important those goals are to the mission of the college. Returning to an earlier example, many institutions state a commitment to various moral virtues that they desire for their students including sensitivity to cultures and beliefs different from their own, service to the community, religious commitment, honesty, etc. To use grades as a carrot/stick to achieve these institutional objectives is a serious ethical violation. For example, suppose five students are suspected of cheating on a take-home final exam, the impact of which is that all five will have a course grade of F. It would be wrong for the instructor, department chair,

---

<sup>21</sup> Battaly (2006) provides an interesting treatment of teaching virtues in the philosophy classroom.

or dean to offer one of the students a D in the course if he will testify against the other four. This is not a condemnation of campus plea-bargaining generally, but only of such deals that involve course grades.

The point is that even though course grades will be used instrumentally in the grade marketplace of job-hunting, graduate admissions, etc., course grades must not be constructed instrumentally. Constructing grades to motivate, cajole, or praise students is unethical. Grades may well have such effects causally, but they should not be assigned to students in order to bring about those effects.

Finally, what about consequences of grading that are harmful in some sense? Is there a teaching equivalent of the ancient dictum in medicine, *primum non nocere* (“first, do no harm”)? Obviously, an academic evaluation process that causes physical pain is ethically suspect—think of not allowing students to use the restroom during a three-hour final exam—while other painful evaluations are likely to be permissible, e.g., a long and difficult piano recital that causes pain in the hands and forearms.

On the other hand, if a grading practice causes psychological harm to a student, is it *prima facie* unethical? Some psychological discomfort often accompanies testing, for example. This essay is devoted to weeding out unjust grading practices precisely because injustice is a harm. But not all harms that are a consequence of grading are unjust and Principle 2.2 states that consequences of grading are generally not ethically relevant anymore than my unhappiness caused by my hypertension diagnosis is medically relevant to the physician’s pronouncement that I have high blood pressure. Unfortunately, the issue of psychological harm in education is a complex social issue. The prevention of psychological harm underlies the K-12 practice of social promotion, for example, and can be found in many other pre-college educational practices, where developmental and political concerns are believed to outweigh the informational value of simple honesty. In cases of genuine medical handicaps, accommodations can be made as outlined above, but *primum non nocere*, referring to mild psychological distress, for example, does not appear to apply to university grading.

## **Punitive Grading**

A possible exception to Principle 2 (grading should be relevant to the academic objectives of the course) is punitive grading for academic dishonesty. First, what is punitive grading? Let’s assume that a grade is punitive if it is lower than the instructor’s best estimate of the student’s knowledge or competence in the course content would otherwise determine. Consider the following example. If Sally cheats on the midterm exam by having someone else take the exam in her place, then the instructor would rightly grade Sally’s exam with a zero since she has failed to demonstrate her knowledge of course content. The zero may have no effect on Sally’s course grade, it might lower the course grade by a letter, etc. It might even lower Sally’s course grade to an F. Still, any such consequent lowering of the grade would not be punitive, *per se*. However, given the same circumstances, if the instructor assigns Sally an F in the course as a matter of department policy in cases of cheating on an exam, regardless of Sally’s level of properly demonstrated competence in the subject matter, then the F grade is punitive. Is such a grade fair?

When I first began working on the grading issue in the late 1980s, I supported punitive grading for academic dishonesty and I contributed to a college policy that allowed individual instructors to automatically fail a student in the course if the student engaged in plagiarism. But, we’ve already seen that to view grading as a reward/punishment system is fundamentally wrong-headed. And, we also know that what misleads us in our thinking about grades is that grades are directly tied to economic rewards and punishments in the form of scholarships, graduate admissions, social status, academic probation, academic dismissal, and so on.

---

Consequently, we should rethink our concept of grades rather than developing punitive grading policies, both at the individual level and the institutional level. This clearly includes cases of academic dishonesty, but it also includes grade reductions based on attendance. If I lower a B student's course grade to a C because she missed more than my magic number of unexcused cuts, I have corrupted the grade itself for all future transcript readers. This is surely a serious violation of professional responsibility. Additionally, I have treated the student unfairly because my grade is a lie, literally a libel. The student is a B student, not a C student, and I have misrepresented her level of competence in the course material to every reader of her transcript.

I certainly believe that academic dishonesty should be punished, sometimes very severely, but I am no longer convinced that the punishment should be in the form of a lowered or failing course grade (recalling that a zero grade on the dishonestly completed grade component is not punishment, per se). Punishment for academic dishonesty could include formal expulsion from the class, or even from the college, accompanied by a WD or a WF mark on the transcript. Harvard's one-strike-and-you're-out expulsion policy is an example here. But to drop an A student to a C because she allowed a friend to copy from her exam is a clear corruption of the grade itself. The A student remains as such irrespective of her dishonorable academic conduct. As with poor attendance, tardiness, and other forms of poor comportment, there are many methods by which we can seek to deter bad conduct—and at the same time, convey that comportment information to those who require it—without inserting comportment into the course grade. We can reprimand the student, expel the student, suspend the student for a semester, remove the student from campus housing, bar participation in athletics, or even fine the student. There are many non-grade sanctions available. And again, the counterargument that such sanctions are too difficult to administer, too expensive, too politically unpleasant, etc., are not persuasive reasons for avoiding the right thing, particularly when doing what is right is so simple.

## V. Applications

### Measuring Fairness

How can we measure the fairness of a course component such as a quiz or an exam? There is a substantial literature on statistical analysis of grading instruments, the normalization of those instruments, establishing test-retest reliability, and demonstrating validity of grade components.<sup>22</sup> One might think that such corrective techniques apply only to mechanically graded objective tests, but any set of numerical scores can be statistically evaluated, even grade components such as essay quizzes and papers. (This is an argument for abandoning the use of letter grades in evaluating grade components.)

Regardless of the nature of the grade component, a rough but useful criterion of fairness is the class mean, or average (or other measurements of central tendency). In a normal distribution of student scores, the class mean of a fair exam will fall in the middle of the C range of the professor's grading scale. For a professor who grades on a 60 pass scale, the mean score on a fair exam will be 75. If the class mean is higher than the midpoint of the C range, the exam may have been too easy (or the students were unusually well-prepared). Likewise, if the class mean is lower than the midpoint of the C range, the exam may have been unfairly difficult (or the students were unusually careless or poorly prepared). Deviation from the mean is also relevant here, though I find an actual calculation of variance or standard deviation

---

<sup>22</sup> For example, see Winters 2002.

does not provide much more information than can be had by simply observing that there is little dispersion or a lot of dispersion of scores. If I find that no one scores above a C on the midterm exam—not even the strongest students—a defective exam is a more likely cause than is universal lack of preparation.

Assuming again that the grade distribution is normal—it's often not—such situations may require recentering the scores, or what is often called “curving.” “Curving,” in ordinary use, is nothing more than adding or subtracting to or from scores on an exam in order to force the class mean to the midpoint of the C range. Some professors compute curved grades with a calculator or spreadsheet program while instructors with good estimation skills may “eyeball” the scores and add what is estimated to be necessary. (A more complex form of curving can be accomplished by regrading a portion of the exams and then writing a function that maps all of the original scores onto points in the new distribution, but that requires more math than many faculty will want to employ.) A few, perhaps a very few, professors may curve grades down, but this is likely to be uncommon since the adjustment penalizes students who did well on the exam.

On the other hand, “grading on the curve” refers to the quite different process of treating a set of grades as being distributed along the bell-shaped curve known in statistics as the normal curve. If a set of grades does actually form a normal distribution, there will be a small percentage of As and Fs, a larger percentage of Bs and Ds, with the largest fraction of grades falling into the C range. Relatively few professors use such a method of grading (see Principle 1.3 above). In skill-oriented courses such as conversational language, logic, mathematics, computer programming, musical performance, etc., grades are rarely normally distributed. The same is true of most upper division courses. Students tend to do comparatively well in their upper division major courses at least in part because weak students are no longer in the major by that point in their education.

### **Grade Component Weights and Sample Size**

Suppose that my course design includes a midterm and a final exam of equal weight. The midterm might have 100 raw points while the final exam has 300 raw points. Some students complain that such an arrangement is unfair. Are they right? No, the students are mistaken. Within obvious extremes, the number of raw points on exams of equal value is generally unrelated to the fairness of the exams. The reason for this is simple. All exams are samples of the students' knowledge of the material. All that is essential to fairness here is that there are enough questions to adequately sample the range of knowledge over which you wish to examine your students.

### **Averaging Letter Grades**

A true average of letter graded course components cannot be achieved by the “eyeball” method in any but the most obvious cases. A fair calculation of a course grade requires converting all letter grades to their numeric equivalents. Here's the rub: should a B, for instance, be equated with the minimum or the maximum numeric value in the B range, or somewhere in between? This is a critical decision because it will affect the average. For example, on a “60.0% pass” grading scale, should a “B” paper grade be converted to 80.0%, 85.0%, or 89.99%? Whatever choice is made here must be applied uniformly to the other letter grades.

If pluses and minuses are assigned, the problem remains. Is a D- to be paired with the bottom of the D- range (60.0%), the midpoint of the D- range (61.67%), or the top of the D- range (63.33%)? A few examples with a calculator or spreadsheet will show that these seemingly insignificant differences can

---

result in an entire letter grade difference for a student's course grade. Pairing a letter grade with the bottom of its numeric range generates higher grades than pairings with the top of the range. That is, it's harder to get an A+ if an A+  $\geq$  99.99% than if an A+  $\geq$  96.67%. This is another reason not to use letter grades when evaluating course grade components such as papers, quizzes, debates, seminar presentations, etc.

### **Grade Inflation and Fair Grades**

We applaud when wages increase and when criminals are punished with heavier fines and longer prison terms, yet some faculty and administrators are fearful of either "too many" As or "too many" Fs. Swarthmore College is famous (or infamous) for its long-time reputation of "hard" grading. "Anywhere else, this would have been an A!" say the Swarthmore students. Other elite institutions have been recently reported in the higher education press to be taking a stand against grade inflation, i.e., grades that are disproportionately high relative to the level of the student's knowledge of the course content. And, we can all think of cases on our own campuses where instructors' grade reports have been flagged for review by the Dean because the course failure rate has exceeded some magic number.

There is a substantial literature on the alleged phenomenon of grade inflation, including arguments that there is no grade inflation at all, but rather, higher levels of achievement, etc. While the general topic of grade inflation is beyond the scope of this paper, the issue is connected with concerns of both professional responsibility (truthful reporting) and fairness. But, we can quickly determine that high grades or low grades, per se, are neither irresponsible nor unfair. If the professor's grading is otherwise fair and if most of the students' mastery of the course content is excellent, then that professor should assign mostly As. Similarly, if the professor's grading is otherwise fair and if most of the students have mastered little or none of the course content, then that professor should assign mostly Fs. This point is directly related to Principles 1.3 and 2 above. There is nothing inherently unprofessional or unfair about any particular distribution of grades in a given course.<sup>23</sup>

## **VI. Conclusion**

Fair and ethical grading is based on two fundamental principles. The first principle is that a fair grade is impartial. The second principle is that a fair grade is an expert evaluation of a student's mastery of course content. These two principles and their corollaries identify as unfair and violative of professional duties common practices such as exempting students from completing certain grade components, and grading on attendance, class rank, comportment, effort, institutional values, and moral virtues such as cheerfulness and helpfulness. The two principles of fair grading can be surrendered only by corrupting the informational value of the grade for transcript readers, including students, faculty advisors, employers, awards committees, graduate admissions committees, and others who depend on the transcribed grades as a source of expert information about the student's level of knowledge in various content areas.<sup>24</sup>

---

<sup>23</sup> This shows immediately that grade distributions should never be used as a metric of professional competence.

<sup>24</sup> An earlier version of this paper was presented as the Presidential Address to the American Association of Philosophy Teachers at the *15th Biennial Workshop/Conference on Teaching Philosophy*,

---

University of Toledo, Toledo, Ohio, 7 August 2004. I benefited from the comments I received there, some of which have been incorporated into this paper. The paper is based on some ideas presented under the same title at the Heidelberg College Faculty Research Symposium, Tiffin, Ohio, 8 February 2001. I appreciate the comments of the editor and the anonymous reviewers.

---

Works Referenced

- American Association of University Professors. 2001. *Policy Documents & Reports*, ninth ed. Washington, D.C.: American Association of University Professors.
- Battaly, Heather. 2006. "Teaching Intellectual Virtues: Applying Virtue Epistemology in the Classroom." *Teaching Philosophy* 29, no. 3 (Sept): 191-222.
- Cahn, Steven M. 1986. *Saints and Scamps: Ethics in Academia*. Lanham, Md.: Rowman & Littlefield, 107.
- Campbell, James. 1988. "Grading Philosophy Papers." *AAPT News* 11, no. 1 (Feb.): 5-8. American Association of Philosophy Teachers, California State University at Long Beach, Long Beach, Calif.
- Chartier, Gary. 2003. "Truth-Telling, Incommensurability, and the Ethics of Grading." *Brigham Young University Education and Law Journal* 2003, no. 1: 37-82.
- Cross, Lawrence H. 1995. "Grading Students." *ERIC/AE Digest*. Washington, D.C.: Clearinghouse on Assessment and Evaluation, The Catholic University of America, Fall, <http://ericae.net/ericdb/ED398239.htm> (accessed March 16, 2009).
- Curren, Randall R. 1995. "Coercion and the Ethics of Grading and Testing." *Educational Theory* 45, issue 4 (Fall): 425.
- Curren, Randall R., ed. 2007. *Philosophy of Education: An Anthology*. Malden, Mass.: Blackwell Publishing.
- Davis, Barbara Gross. 1993. *Tools for Teaching*. San Francisco: Jossey-Bass.
- Dayton, John and Anne Proffitt Dupre. 2005. "Grading Questions You Were Afraid to Ask, Answers You Need to Know." <http://www.mcgeorge.edu/documents/agencies/dayton%20paper%20ser.doc> (accessed March 16, 2009).
- Fenner, David E. W., ed. 1999. *Ethics in Education*. New York: Garland Publishing.
- Freeman, Mike. 2003. "When Values Collide: Clarett Got Unusual Aid in Ohio State Class." *New York Times*, 13 July, <http://query.nytimes.com/gst/fullpage.html?res=9407E3D81F3DF930A25754C0A9659C8B63&sec=&spon=&pagewanted=all> (accessed March 16, 2009).
- Friedman, Stephen J. 1998. "Grading Teachers' Grading Policies." *NASSP Bulletin* 82, no. 597 (April): 77.
- Goldman, Alan H. 1980. *The Moral Foundations of Professional Ethics*. Totowa, N.J.: Rowman and Littlefield, 287.

- 
- Gullickson, Arlen R. and The Joint Committee on Standards for Educational Evaluation. 2003. *The Student Evaluation Standards: How to Improve Evaluations of Students*. Thousand Oaks, Calif.: Corwin Press.
- Hammons, J. O. and J. R. Barnsley. 1992. "Everything You Need to Know About Developing a Grading Plan for Your Course (Well, Almost)." *Journal on Excellence in College Teaching* 3: 51-68.
- Howe, Kenneth R. 1988. "An Evaluation Primer for Philosophy Teachers." *Teaching Philosophy* 11, no. 4 (December): 315-328.
- Kant, Immanuel. [1887], 1796. *The Philosophy of Law: An Exposition of the Fundamental Principles of Jurisprudence as the Science of Right*. Translated by William Hastie. Edinburgh: T. and T. Clark, <http://oll.libertyfund.org/title/359> (accessed March 16, 2009).
- Keith-Spiegel, P., A. F. Wittig, D. V. Perkins, D. W. Balogh, and B. E. Whitley, Jr. 1993. *The Ethics of Teaching: A Casebook*. Muncie, Ind.: Ball State University Press.
- Keen v. Penson, 970 F.2d 252, 256 (1992).
- Keith-Spiegel, P., B. G. Tabachnick, and M. Allen. 1993. "Ethics in Academia: Students' Views of Professors' Actions." *Ethics and Behavior* 3: 149-162.
- Milton, Ohmer, H. R. Pollio and J. A. Eison. 1986. *Making Sense of College Grades: Why the Grading System Does Not Work and What Can Be Done about It*. Ann Arbor, Mich.: Proquest Information and Learning.
- Milton, Ohmer. 1992. "We Must Think Anew." *Journal on Excellence in College Teaching* 3: 19-32.
- Moll, Marita. 1999. "The History of Grading in Three Minutes." *The Learning Team* 2, no. 3 (Spring), <http://www.teachers.ab.ca/Quick%20Links/Publications/The%20Learning%20Team/Volume%202/Number%203/Pages/The%20history%20of%20grading%20in%20three%20minutes.aspx> (accessed March 16, 2009).
- Office of the Dean of the College. 2005. "Princeton University Grading Policies In Undergraduate Courses And Independent Work." Princeton: Princeton University, 21 January, <http://www.princeton.edu/odoc/faculty/grading/> (accessed March 16, 2009).
- Panza, Chris. 2007. "Grading on Comportment." *In Socrates' Wake: A philosophy teaching blog*: 26 August, <http://insocrateswake.blogspot.com/2007/08/grading-on-comportment.html> (accessed March 16, 2009).
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

- 
- Rodabaugh, R. C. 1996. "Institutional Commitment to Fairness in College Teaching." In L. Fisch, ed., *Ethical Dimensions of College and University Teaching*. San Francisco: Jossey-Bass, 37-45.
- Sabini, John and John Monterosso. 2003. "Moralization of College Grading: Performance, Effort, and Moral Worth." *Basic & Applied Social Psychology* 25, no. 3: 189.
- Schrag, Francis. 2001. "From Here to Equality: Grading Policies for Egalitarians." *Educational Theory* 51: 63.
- Scriven, Michael. 1967. "The Methodology of Evaluation." In *Perspectives on Curriculum Evaluation* (AERA Monograph Series—Curriculum Evaluation), eds. Ralph Tyler, Robert Gagné and Michael Scriven. Chicago: Rand McNally. Widely reprinted, e.g., in B. Worthen and J. Saunders, eds., *Educational Evaluation: Theory and Practice*, 60-103. Belmont, CA: Wadsworth Publishing, 1973.
- Smith v. School City of Hobart, 811 F. Supp. 391, 397-98 (1993).
- Terwilliger, James. 1977. "Assigning Grades: Philosophical Issues and Practical Recommendations." *Journal of Research and Development in Education* 10: 21.
- Urmson, J. O. 1950. "On Grading." *Mind*. Reprinted in Anthony Flew, ed. with Introductions. [1965], 1951, 1953. *Logic and Language* (First and Second Series). Garden City, N.Y.: Anchor Press/Doubleday.
- Weis, Gregory F. 1995. "Grading." *Teaching Philosophy* 18, no. 1 (March): 3-13.
- Weller, L. David. 1983. "The Grading Nemesis: An Historical Overview and a Current Look at Pass/Fail Grading." *Journal of Research and Development in Education* 17, no. 1 (Fall): 39-45.
- Whitley, Jr., Bernard E., David V. Perkins, Deborah Ware Balogh, Patricia Keith-Spiegel, and Arno F. Wittig. 2000. "Fairness in the Classroom." *APS Observer* 13, no. 6, [http://www.psychologicalscience.org/teaching/tips/tips\\_0700.cfm](http://www.psychologicalscience.org/teaching/tips/tips_0700.cfm) (accessed March 16, 2009).
- Wilson, Scott D. 2006. "Peer-Review Assignments." *Teaching Philosophy* 29, no. 4 (December): 327-342.
- Winters, R. Scott. 2002. "Score Normalization as a Fair Grading Practice." *ERIC Digest* (December). ERIC Identifier: ED470592. College Park, Md.: ERIC Clearinghouse on Assessment and Evaluation, [http://permanent.access.gpo.gov/websites/eric.ed.gov/ERIC\\_Digests/ed470592.htm](http://permanent.access.gpo.gov/websites/eric.ed.gov/ERIC_Digests/ed470592.htm) (accessed March 16, 2009).
- Woozley, A. D. 1973. "Injustice." *American Philosophical Quarterly Monograph Series: Studies in Ethics* 7.

Zak, F. and C. C. Weaver, eds. 1998. *The Theory and Practice of Grading: Problems and Possibilities*. Albany, N.Y.: State University of New York Press.

*Daryl Close, Department of Philosophy, Heidelberg University, Tiffin, Ohio 44883  
dclose@heidelberg.edu*