Appendix: The Uses and Abuses of Grades and Examinations*

STEVEN M. CAHN (Ph.D., Columbia University) is professor of philosophy and chairman of the department at the University of Vermont. Previously he taught at Dartmouth College, Vassar College, the University of Rochester, and New York University. He is the author of The Eclipse of Excellence: A Critique of American Higher Education; A New Introduction to Philosophy; and Fate, Logic, and Time. Among the books he has edited are The Philosophical Foundations of Education; Philosophy of Religion; and Classics of Western Philosophy. His articles have appeared in such publications as The Encyclopedia of Philosophy, The Journal of Philosophy, the American Philosophical Quarterly, The New York Times, and The Chronicle of Higher Education.

^{*} Although this essay originally was written for this volume, it appeared first as part of the author's work *The Eclipse of Excellence: A Critique of American Higher Education* (Washington, D.C.: Public Affairs Press, 1973). Reprinted here by permission of Public Affairs Press.

It is astonishing to realize how little a college teacher may know about the academic lives of his students, and conversely, how little his students may know about the academic life of their teacher. I recall one professor who taught a large lecture course for several years without ever realizing he was addressing a captive audience, since unbeknownst to him, the course was required for graduation. On the other hand, I have spoken to students unaware that just as a student can be required to take a course he would prefer not to take, so a teacher can be required to teach a course he would prefer not to teach. How many teachers, even in a very small class, know whether their students are sophomores or seniors, a matter of some importance to the students? But again, how many students know whether their teacher is an assistant or full professor, a matter of some importance to the teacher? It may never occur to a teacher that the sleepy students in his 8 A.M. class are there only because all other sections of the course were already closed when they registered. But, likewise, it may never occur to the students in the 8 A.M. class that their teacher is standing wearily before them at such an hour only because he lacks the seniority to claim any other time.

Such mutual ignorance extends over many aspects of academic life and is nowhere more apparent than in matters regarding examinations and grades. A basic source of the misunderstandings which surround evaluations of student work lies in the fact that normally such evaluation has vital consequences for the one being evaluated, whereas it has no such consequences for the one who does the evaluating. The grades a student receives not only determine whether he graduates with honors or fails out of school; they may also guide him in choosing his field of specialization, affect his plans for graduate study, and ultimately influence his choice of career. On the other hand, the grades a teacher gives do not affect his professional stature, his commitment to a field of study, or his future success as a scholar. A student may for a long time harbor a deep resentment against a teacher who grades him harshly, but were he to confront that teacher years later, the teacher might not even remember the student and would almost surely not remember the grade. Indeed, the teacher would most probably be astounded to learn the student cared so deeply about the grade. I once heard a woman who had taught for over thirty years remark in a faculty meeting that she could not understand why students were so interested in grades. Apparently in

moving from one side of the desk to the other she had developed amnesia.

Some students believe that teachers are fond of examinations and grades, that they employ these devices in order to retain power over the students. But although undoubtedly a few teachers do possess such motives, most do not. A scholar enjoys reading and writing books, not making up questions to test the knowledge others may possess. And a great many more fascinating things can be found to do than read one hundred or so answers to the same question and try to decide how many points each answer is worth. Whether Johnny understands the problem of induction is not crucial to Professor Smith's intellectual life, for Professor Smith finds the problem highly stimulating, even if Johnny neglects to study it.

The system of examinations and grades thus places important decisions affecting students' lives in the hands of those who are comparatively unaffected by these decisions and perhaps quite uninterested in making them. Such a situation is fraught with unpleasant possibilities, these often compounded by the difficulty of constructing and applying suitable examination and grading procedures. But to refer to "suitable examination and grading procedures" implies that such procedures are intended to fulfill certain worthwhile purposes, and so we would do well to consider just what those purposes are. Why bother with examinations or grades at all?

Examinations ideally serve at least four significant purposes. First, an examination provides the opportunity for a student to discover the scope and depth of his knowledge. Much like an athlete who tests himself under game conditions or like a violinist who tests himself under concert conditions, a student tests himself under examination conditions and thereby determines whether he is in complete control of certain material or whether he possesses merely a tenuous grasp of it. It is one thing to speak glibly about a subject; it is something else to answer specific questions about that subject, relying solely upon one's own knowledge and committing answers to paper so they can be scrutinized by experts in the field. A proper examination procedure makes clear to the student what he knows and what he does not know and thus can serve as a valuable guide to further study. By paying close attention to the results of his examination, a student can become aware of his strengths and weaknesses. He can learn whether his methods of study are effective, and he can recognize the areas of

a subject in which he needs to concentrate his future efforts. In short, an examination enables a student to find out how well he is doing and assists him in deciding how he can do better.

Students, however, are not the only ones who are tested by an examination, for the second purpose examinations should serve is to provide an opportunity for a teacher to discover how effective his teaching has been. By carefully analyzing his students' examination papers, a teacher can learn in what ways he has succeeded and in what ways he has failed. Of course, many teachers would prefer to believe the reason three-quarters of their students missed a particular question is that the students are not bright or have not studied hard enough. But in this matter, college teachers have something to learn from those who teach in elementary school. When three-quarters of an ordinary third-grade class find multiplication confusing, the teacher does not assume the students are not bright or have not studied hard enough. He assumes his teaching methods are in need of improvement. A college teacher ought to arrive at the same conclusion when three-quarters of his class are confused by a fundamental point he thought he had explained clearly. In one sense, then, teachers as well as students can pass or fail examinations, for by paying close attention to the results of his students' efforts, a teacher can become aware of the strengths and weaknesses of his instruction. He can learn whether his methods are effective, and he can recognize the areas of a subject in which he needs to concentrate his future efforts. In short, an examination enables a teacher to find out how well he is doing and assists him in deciding how he can do better.

We have thus far considered examinations only as tests of learning, but they can be more than a means of evaluating previous learning experiences: they can be themselves worthwhile learning experiences. During an examination most students are working with an extraordinarily high degree of concentration. If the examination questions place familiar material in a slightly unfamiliar light and thereby lead students to develop for themselves significant connections between various aspects of the subject matter, then the students will be working intensely on challenging, important problems and so gain intellectual perspective. Ironically, in this day of large lecture classes, examinations sometimes provide greater opportunity for active learning than any other part of the course. It is not unusual to hear student complaints about uninspired, unrewarding examinations. Such complaints are entirely legitimate, for a boring, banal examination indicates pedagogic laziness and is a waste of a potentially valuable learning experience. Long after completing a course, students who have forgotten virtually everything else may still remember some of the examination questions. They should be worth remembering.

An examination, however, consists of more than the two or three hours spent sitting in the examination room. Most students prepare for examinations, and such preparation itself possesses significant educational value. The nature of an examination requires that one not know what questions will be asked or which aspects of the subject matter spotlighted. The only adequate preparation for an examination is a thorough study of all the subject matter and a careful consideration of as many as possible of its various interconnections. In trying to anticipate the examination questions, a student is led to analyze and synthesize the course material, thereby strengthening and solidifying his grasp of the subject matter.

In this connection it is important to note that the writing of a term paper, though potentially a beneficial educational experience, is not a suitable substitute for preparing for an examination. In writing a term paper, even one which is given a strict time limit and misleadingly dubbed "a take-home examination," a student need only master those parts of the course material bearing directly on his topic. Rarely does a term paper require mastery of most or even very much of the course material. Furthermore, it is not difficult to copy ideas from a book, alter them slightly so as to avoid the charge of plagiarism, and use them in a term paper without ever thoroughly understanding them. Such a tactic is almost impossible in an examination, for few students have a strong enough memory to answer questions intelligently without understanding their answers. Thus, preparing for an examination is in some ways, though not all, more demanding and more rewarding than writing a term paper.

This fact was strikingly brought to my attention several years ago by a student who came to see me after I had returned her examination paper. She had received a C and was very disappointed, for, as she explained, she had always been an A student. I asked her whether she had studied as hard for this examination as for previous ones, and to my surprise she informed me that never before in her academic career had she taken an examination. As it turned out, she had gone to a "progressive" secondary school where examinations were considered outmoded, and she had then attended a college that prided itself on having replaced all examinations with term papers. I was fascinated by this woman's academic background and inquired whether she thought she had been helped or hindered by it. She replied that until she had taken this examination she had always assumed it was to her advantage to have avoided the pressure of examinations, but that now she believed her grasp of previous course material rather flimsy. She had learned how to write term papers but never had thoroughly mastered an entire body of material so that she could draw upon it at will and utilize it effectively wherever it was called for. In short, she had never received the benefits of preparing for an examination.

Of course, examinations serve yet another purpose, for they are in part the basis on which course grades are determined. However, since we have already seen that examinations provide an opportunity to discover the scope and depth of a student's knowledge, we have little reason to doubt that if grades are to be given, they should be based, at least to some extent, on the results of examinations. The crucial question is: why should grades be given?

Ideally, a grade represents an expert's opinion of the quality of a student's work within a specified area of inquiry. Viewed in this perspective, a grade serves a variety of significant educational purposes. First, it is to a student's advantage to be aware of his level of achievement, for that information can be a valuable aid to him in assessing his past efforts, evaluating his present abilities, and formulating his future plans. Knowing whether one's approach to a subject has been fruitful is a helpful guide toward further study; recognizing one's strengths and weaknesses is vital to intellectual growth as well as to decisions regarding how one's abilities might most effectively be utilized in and out of school. A college student is directly concerned with questions such as "Which courses should I take?", "Which fields should I specialize in?", "Which graduate schools, if any, should I apply to?", and "Which career should I choose?" Intelligent answers to all these questions depend, among other factors, upon the individual's academic abilities and accomplishments, and he can measure these reliably, though not infallibly, by his grades. Granted a teacher's judgment may occasionally be mistaken, at least his judgment is based upon relevant expertise and experience and is not subject to the sort of delusions which so often distort self-evaluation. A student may not always be pleased by the knowledge grades afford,

but the important point is that such knowledge is almost always useful to him.

Students, though, are not the only ones to whom such knowledge is useful, for in order for a teacher to provide the detailed educational advice often so helpful to a student, he needs to have an exact record of the student's academic performance. How can a teacher intelligently advise a student in choosing his program of study and in planning for the years after graduation if an accurate measure of the student's level of achievement is unavailable? If, for example, a chemistry teacher does not know how well a student has done in his various science and mathematics courses, how can the teacher intelligently advise the student which level of chemistry to study, which areas in the student's background need strengthening, and whether it is reasonable for the student to continue work in graduate school? And if the student should decide to become a political science major, how can a teacher in that discipline intelligently advise the student what course of study to follow without knowing his level of achievement in history, economics, sociology, philosophy, and nowadays even in mathematics? In short, students' academic records are a great aid to those teachers who try to use their knowledge and experience to advise students wisely. But if a student's record is sketchy, vague, and inadequate, the advice he receives will most likely also be sketchy, vague and inadequate.

We have already noted that grades can be a valuable guide to a student in planning for the years following his graduation, but we should note as well that grades are a valuable guide to those who must make critical decisions directly affecting a student's future plans. Graduate work usually presupposes a firm command of undergraduate work, and thus most graduate schools necessarily employ selective admission policies. Those who face the difficult task of deciding whether a particular student is to be admitted to graduate school can make that decision intelligently only if they are aware of the student's level of achievement in his various college courses, and grades are a reliable, though not infallible, measure of such achievement.

On occasion, however, it is proposed that instead of receiving an applicant's grades a graduate admissions committee receive instead recommendations written by each of the teachers with whom the applicant has studied. But this proposal is impractical and, even if feasible, would nevertheless be inadvisable. The proposal is impractical for at least two reasons. First, the members of an admissions committee do not have the time to read twenty-five or thirty letters about each applicant. In the case of some of the larger graduate schools, an admissions committee with twenty-five letters for each applicant would be facing more than twenty-five thousand letters and could not possibly be expected to spend the time necessary to do justice to that amount of material. Second, the large size of so many college classes makes it virtually impossible for a teacher to know each of his students personally. Thus he would be reduced to writing such conventional comments as "DeWitt is an excellent student who has mastered all of the course material" or "Davis is a fair student who has mastered some, though not much of the course material." But what do these comments mean except that De-Witt did A work in the course and Davis did C work?

However, even if it were feasible for every one of a student's teachers to write a personalized comment about him and for an admissions committee to read all of these comments, still they would not be an adequate replacement for grades. Recommendations sometimes contain valuable information, but taken by themselves they are often difficult to evaluate. A remark one teacher considers high praise may be used indiscriminately by another, and a comment employed by one teacher to express mild commendation may be used by another teacher to express mild criticism.' Furthermore, many recommendations are hopelessly vague and tell more about the teacher's literary style than about the academic accomplishment of the student. Thus although letters of recommendation may be helpful when used in conjunction with grades, alone they are no substitute for the relatively standardized measure of achievement grades effectively provide.

Such a standardized measure of achievement also affords a reasonable basis upon which to decide whether a student ought to be permitted to continue in school, whether he ought to be granted a college degree, and whether he ought to be awarded academic honors. These decisions, however, have all been the subject of controversy, and so we would do well to consider each of them separately.

A student who consistently does unsatisfactory work is squandering the resources of his college, wasting the time and energy of his teachers, and failing to contribute to, perhaps even interfering with, the education of his classmates. Such a student does not belong in the school he is attending, and, for the benefit of all concerned, should be asked to leave. But which students are doing unsatisfactory work? In answering this question it is clearly most sensible to rely upon the expert judgment of the faculty, and their judgment, as noted previously, is reliably reflected by a student's grades.

The faculty's expertise ought also to be relied upon in deciding whether the quality of a student's work justifies his being granted a college degree. Because most students are charged tuition fees, it is tempting to conceive of a college as an educational store in which the student customers pay their money and are then entitled to a degree. But a college degree is not purchased; it is earned. It represents to the community the college's certification of a student's academic achievement, certification respected because it is backed by the expertise of the faculty. If every student who paid his tuition automatically received a degree, or if degrees were awarded by the vote of the student body, then they would become educationally meaningless and functionally worthless. In order for a college degree to retain its value and for a college education to retain its significance, the granting of degrees must be based solely upon substantial academic achievement as evaluated by recognized experts. The experts are the faculty, and their evaluations are indicated by the grades they give.

Grades also provide an effective means of determining which students are deserving of academic honors. Such honors are both an added incentive for students to pursue their work diligently and a symbol of a college's commitment to academic excellence. But in order for honors to possess such significance, they must not be granted indiscriminately or on the basis of a student's popularity. Rather, they must be awarded only to those who have attained a high level of scholarly achievement. And grades provide a standardized measure of such achievement.

Grades serve one final purpose: to motivate students to study. In the classroom, as in most areas of life, those who expect their work to be evaluated tend to do that work more assiduously. Without grades, many students might possess sufficient interest to casually peruse the course material, but few would be strongly enough concerned to devote themselves to the mastery of that material. Of course, there are a handful of students who would thoroughly study all of their course material even if they did not receive any grades. These are the saints of the academic world. But a teacher should no more assume all his students saints, than he should assume all his neighbors saints. In both cases he would do well to hope for the best but prepare for the worst. What should be remembered is that grades have helped many students who otherwise would have neglected their work, and have led some to discover for themselves the intrinsic joys of scholarship.

We must recognize, however, that notwithstanding the many worthwhile purposes examinations and grades are intended to fulfill, much criticism has been directed against these educational tools. It has been claimed that examinations fail to provide a sound basis for evaluating a student's achievement but, instead, have the effect of inhibiting his independence and stifling his creativity. It has also been claimed that grades are inherently inaccurate devices which, in attempting to measure people, succeed only in traumatizing and dehumanizing them. These charges are certainly serious, and each of them ought to be examined in detail.

Consider first the claim that examinations do not provide a sound basis for evaluating a student's achievement. Those who defend this claim argue that examinations require a student to demonstrate his knowledge under adverse conditions; he must answer a restricted set of questions within a limited amount of time, and the implicit pressure prevents many from doing their best work. Thus the results of examinations are said to be invalid.

But this line of argument overlooks the vital consideration that although examinations put pressure on students, such pressure exists whenever an individual attempts to prove to experts his competence in their field. For instance, an athlete feels pressure when he tries out for a professional team; likewise a violinist when he auditions for an orchestral position. Pressure is inherent in such situations, for experts have high standards difficult to meet, and one must be able to meet those standards at an appointed time. The ballplayer who appears skillful in practice but plays poorly in league games lacks effective control of the requisite skills. Similarly, the student who sounds knowledgeable in conversation but performs poorly under examination conditions lacks effective control of the requisite knowledge. Thus the pressure of examinations does not invalidate the results of examinations; quite to the contrary, if there were no such pressure, the examination process would be amiss.

A second criticism of examinations is that they inhibit a student's independence, that they discourage him from pursuing topics of interest to him and instead force him to study topics of interest to his teacher. Thus, it is said, examinations impede rather than promote the learning process.

This criticism, however, rests upon the mistaken assumption that learning a particular subject matter involves nothing more than learning those aspects of the subject matter one happens to find interesting. For example, to attain a thorough knowledge of American history, it is not sufficient to learn the history of the American Indian, no matter how interested one may be in the Indians, for American history, like any significant area of inquiry, has many important aspects, all of which must be mastered in order to attain a thorough knowledge of the field. But who is to decide which aspects of a subject matter are most important? The teacher is the recognized expert, and so he is in a position to make intelligent curricular decisions. Furthermore, the teacher's responsibility is to use his expertise to further a student's education, to guide him in studying important aspects of the subject matter he might otherwise neglect. Such guidance, in one sense, interferes with a student's independence, but in another, more significant, sense, liberates him from his own narrow preoccupations and leads him to less restricted, more independent thinking. And that, after all, is one of the essential purposes of a liberal education.

Another criticism of examinations is that they stifle a student's creativity, that they emphasize the mindless reiteration of facts and techniques instead of encouraging original, imaginative thinking about significant issues. Thus, again it is said, examinations impede rather than promote the learning process.

But this criticism is mistaken for at least two reasons. First, not all examinations emphasize learning by rote, only poor examinations. Good examinations, as pointed out previously, place familiar material in a slightly unfamiliar light, so that in preparing for and taking examinations, students are led to develop for themselves significant connections between various aspects of the subject matter. Of course, an examination does not normally require the same degree of original, imaginative thinking required by a demanding term paper topic. But it must be remembered a term paper does not require mastery of most or even very much of the course material; only examinations do. In other words, the two tasks serve different purposes, and there is no point in criticizing one for not fulfilling the purposes of the other.

The second reason why the criticism in question is mistaken is that it overlooks that in order to master any significant field of inquiry, one must acquire secure control of certain fundamental information and skills. As Whitehead wrote, "There is no getting away from the fact that things have been found out, and that to be effective in the modern world you must have a core of definite acquirement of the best practice. To write poetry you must study metre: and to build bridges you must be learned in the strength of material. Even the Hebrew prophets had learned to write, probably in those days requiring no mean effort. The untutored art of genius is-in the words of the Prayer Book-a vain thing, fondly invented."² It is simply unrealistic to suppose that original, imaginative thinking of a sustained and productive sort flows from the minds of those ignorant of the fundamental information and skills related to their field of inquiry. Of course, it has been said that the mark of a knowledgeable person is not what he knows, but whether he is adept at looking up what he needs to know. But if this were so, then the most knowledgeable people in the world would be librarians. The fact is that a person who lacks fundamental information and skills is not in a position to understand and intelligently evaluate material confronting him, so he is unable to connect ideas in the ways necessary for sustained, productive thinking. And even if, as is highly doubtful, such an individual had the time to research everything he needed to know, he would not know what to research, for he would not be aware of all he needed to know. But how can it be determined whether an individual possesses the fundamental information and skills related to his field of inquiry? Examinations enable both teacher and student to make such determinations effectively, and thus, rather than stifling creativity, help to provide the framework within which original, imaginative thinking can be most productive.

Turning now from criticisms of examinations to criticisms of grades, consider first the claim that grades are inherently inaccurate. Those who defend this position argue that the same paper would be graded differently by different instructors, and therefore a student's grade is not a reliable measure of his achievement but merely indicates the particular bias of his instructor.

However, a student's work is generally not judged with significant difference by different instructors. In fact, teachers in the same discipline usually agree as to which students are doing outstanding work, which are doing good work, which are doing fair work, which are doing poor work, and which are doing unsatisfactory work (or no work at all).3 Of course, two competent instructors may offer divergent evaluations of the same piece of work. But the fact that experts sometimes disagree is not, of course, reason to assume there is no such thing as expertise. For example, two competent doctors may offer divergent diagnoses of the same condition, but their disagreement does not imply that doctors' diagnoses are in general biased and unreliable. Similarly, two competent art critics may offer divergent evaluations of the same work of art, but such a disagreement does not imply that a critic's evaluations are usually biased and unreliable. Inevitably, experts, like all human beings, will sometimes disagree about complex judgments, but we would be foolish to allow such disagreements to obscure the obvious fact that in any established field of inquiry some individuals are knowledgeable and others are not. And clearly the opinions of those who are knowledgeable are the most reliable measure of an individual's achievement in that field. Thus, although teachers sometimes disagree, they are knowledgeable individuals whose grades represent a reliable measure of a student's level of achievement.

A second criticism of grades is that they traumatize students. Those who support this criticism argue that grades foster competition, arousing a bitterness and hostility which transform an otherwise tranquil academic atmosphere into a pressure-filled, nerve-wracking situation unsuited for genuine learning. In such a situation, it is said, students are more worried about obtaining good grades than about obtaining a good education.

But this criticism emphasizes only the possibly harmful effects of competition while overlooking its beneficial effects. Often only by competing with others do we bring out the best in ourselves. As Gilbert Highet once noted, "It is sad, sometimes, to see a potentially brilliant pupil slouching through his work, sulky and willful, wasting his time and thought on trifles, because he has no real equals in his own class; and it is heartening to see how quickly, when a rival is transferred from another section or enters from another school, the first boy will find a fierce joy in learning and a real purpose in life."⁴ In short, competition fosters excellence, and without that challenge most of us would be satisfied with accomplishing far less than we are capable of.

However, even if competition did not have beneficial effects, it would still be an inherent part of academic life, for it is an inherent part of virtually every aspect of life. Many people have the same goals, but only a comparatively few can achieve them. For example, not everyone who so desires can be a surgeon, a lawyer, an engineer, or a professional football player, and, indeed, marked success in any field of endeavor is necessarily quite rare. Thus competition arises. And since academic success is desired not only for its own sake but also because it relates to success in many other competitive fields, competition will always exist in academic life.

The question then is not whether competition should be eliminated from the academic sphere, but how it can be channelled so as to maximize beneficial effects and minimize potentially harmful effects. The key to this difficult task lies in encouraging each student to strive as vigorously as possible to fulfill his own potential, in praising his efforts when he tries his hardest and in appealing to his sense of pride when his energies flag. Treating him so does not lead him to emphasize good grades rather than a good education, for he cannot achieve a good education without striving for mastery of subject matter. And if grades are awarded as they should be, on the basis of accurate measures of a student's level of achievement, then they will indicate his mastery of subject matter. Thus a student concerned with grades is concerned with a prime component of a good education.

A third criticism of grades is that in attempting to measure people, they succeed only in dehumanizing and categorizing them, depriving them of their uniqueness, and reducing them to a letter of the alphabet. Thus, it is said, grades defeat one of the essential purposes of an education: to aid each individual in developing his individuality.

A grade, however, is not and is not intended to be a measure of a person. It is, rather, a measure of a person's level of achievement in a particular course of study. To give a student a C in an introductory physics course is not to say that the student is a C person with a C personality or C moral character, only that he is a person with a C level of achievement in introductory physics.

Grades no more reduce students to letters than batting averages reduce baseball players to numbers. That Ted Williams had a lifetime batting average of .344 and Joe Garagiola an average of .257 does not mean Williams is a better person than Garagiola, but only that Williams was a better hitter. And why does it dehumanize either man to recognize that one was a better hitter than the other?

Indeed, to recognize an individual's strengths and weaknesses, to know his areas of expertise, his areas of competence, and his areas of ignorance is not to deny but to emphasize his individuality. If Delaney and Delancey are known to their teachers only as two faces in the classroom, then their comparative anonymity is apt to lead to their individual differences being overlooked. But if Delaney has a reputation as an excellent history student with a weakness in mathematics, while Delancey is known as a generally poor student, but one who has a gift for creative writing, then these two students are no longer anonymous cogs in a machine, and their education can be tailored to suit their needs. Thus grades do not dehumanize an individual; on the contrary, they contribute to a recognition of his uniqueness and to the possible development of his individual interests and abilities.

Yet there is one further challenge to the entire system of examinations and grades, for as was pointed out earlier, this system places important decisions affecting students' lives in the hands of those comparatively unaffected by these decisions and perhaps quite uninterested in making them. Such a situation is indeed hazardous, and the potential problems are, of course, compounded by the difficulty of constructing and applying suitable examination and grading procedures. Of course, suitable procedures are the ones most likely to fulfill the worthwhile purposes examinations and grades are intended to serve, and we have already seen what those are. But what specifically are the procedures most likely to fulfill those purposes? And how can it be ensured that teachers will be cognizant of the proper procedures and apply them conscientiously? These are important questions, and they deserve careful consideration.

Constructing a good examination is a creative endeavor, and, as in the case of all creative endeavors, there are no surefire formulas for success; the most one can reasonably hope for are broad guidelines to provide a sound basis for at least partial success. The first such guideline is that an examination should be representative of the course material. Consider, for instance, a course in the history of modern philosophy that devotes two or three weeks to the study of each of five philosophers: Descartes, Leibniz, Berkeley, Hume, and Kant. If the final examination is to serve its proper function as a test of the scope and depth of a student's knowledge of the course ma-

232 / Steven M. Cahn

terial, then the examination should be structured so that a student is called upon to demonstrate considerable knowledge about all five of the authors studied. The examination would be unsatisfactory if it tested only a student's general philosophical ability, not his knowledge of the five authors studied, or if it tested a student's knowledge of only one or two authors studied and permitted him to neglect the others. For whatever such unsatisfactory examinations might be intended to test, they would fail to test adequately the scope and depth of a student's knowledge of the history of modern philosophy.

Of course, an examination representative of the course material need not deny students a choice as to which examination questions they wish to answer. Such a choice is an attractive feature of an examination, since it allows students an opportunity to demonstrate their special interests and abilities. But the crucial point is that such choices should be so arranged that a student's answers will adequately reflect his knowledge of the entire course material. And if certain course material is so essential that all students should be familiar with it, then no choice should be given. For contrary to common practice, students need not always be offered a choice of examination questions. What they should be offered is an examination representative of the course material.

A second guideline for constructing good examinations is posing questions that require detailed answers. Perhaps the most serious fault of college examinations is that they allow a student to talk around the subject matter without ever having to demonstrate more than a superficial knowledge of the course material. Again in contrast to common practice, much can be said in favor of questions that have answers, answers to be found in or at least closely related to the course readings. An examination lacking such questions is not merely a poor test of a student's knowledge but leads him to suppose that thorough knowledge of the course material amounts to no more than knowing a few stray bits of information strung together by some vague generalizations about some even vaguer concepts. Such an examination is worse than no examination at all; it is an educational travesty that leads a student to suppose he has mastered material about which he knows virtually nothing.

But the fact that examination questions ought to require detailed answers is no reason why students should be overwhelmed with truefalse or multiple-choice questions. Though these can sometimes be of educational value, unless they are well-constructed and appropriate to the aims of the course, they turn the examination into a guessinggame that stresses knowledge of minutiae rather than the understanding of fundamental concepts and principles. For instance, only a foolish examination in the history of modern philosophy would be filled with questions such as "The title of Section IX of Hume's An Inquiry Concerning Human Understanding is (a) Of Liberty and Necessity, (b) Of the Reason of Animals, (c) Of Miracles, (d) All of the above, (e) None of the above." On the other hand it would be equally foolish for such an examination to be filled with questions such as "Does it seem to you that anything in the work of Kant helps us to understand ourselves?" What is needed is neither a trivial nor vague question but a sharply defined, significant, challenging question, one such as: "Both Descartes and Berkeley raise doubts about the existence of the material world. Compare and contrast (1) the arguments they use to raise these doubts, and (2) their conclusions concerning the possible resolution of these doubts." An examination with questions such as this not only provides a rigorous test of a student's knowledge but also clearly indicates to the student that mastery of the subject matter is a demanding enterprise, requiring far more intellectual effort than the memorization of trivia or the improvisation of hazy, high-flown vacuities.

If an examination adheres to the two important guidelines just discussed, then there is reason to suppose it will fulfill the worthwhile purposes it should serve. However, several other pitfalls must be avoided in order for an examination to be as effective as possible. First, the examination should not be so long that most students are more worried about finishing than about providing the best possible answers. Of course, if a student takes too long to answer a question, it is clear he does not have secure enough control of the required material. But basically an examination should not be a race against time; it should be constructed so a student working at a normal pace has sufficient time to read the questions carefully, compose his thoughts, write his answers legibly, and reread his work to make corrections. No matter how well constructed examination questions may be, if there is not sufficient time to answer them thoughtfully, the examination will turn into a shambles and be of little use to anyone.

A second pitfall to be avoided is the omission of clear directions at the top of the examination paper. Imagine sitting down to begin work and reading the following directions: "Answer three questions from Part I and two questions from Part II, but do not answer questions 2, 3, and 6 unless you also answer question 9. Question 1 is required, unless you answer questions 3 and 5." By the time a student has fully understood these directions and decided which questions he ought to answer, he will already be short of time.

When a student sits down to take an examination, he is understandably tense and liable to misread the directions, answer the wrong questions, and bungle the examination. If he does so, the fault is probably not his, for the teacher has the responsibility to make the directions so clear that the student will find them virtually impossible to misunderstand. A teacher has sufficient time to work out clear directions, and he owes it to his students to provide such directions. The examination should be a test of a student's knowledge of the course material, not a test of his ability to solve verbal puzzles.

A third pitfall is the failure to inform students of the relative importance of each answer in the grading of the examination. Suppose a student begins work on an examination in which he is required to answer three questions, but is not told the teacher considers the answer to the third question more important than the combined answers to the first two. The student will probably spend an equal amount of time on each, not realizing he should concentrate his time and effort on the third. But his mistake indicates no lack of knowledge on his part. It is simply a result of the teacher's keeping his own intentions a secret. And this secret serves no other function than to distort the results of the examination. It is only fair that a student be informed as to how many points each question is worth, so that he can plan his work accordingly.

One final pitfall must be avoided in order for an examination to fulfill its proper purposes, and this pitfall relates not to the construction of the examination, but to its grading. A teacher is responsible for grading examinations as carefully and fairly as possible. To do otherwise is to waste much of the effort put into constructing and taking the examination, for an examination graded carelessly or unfairly does not provide an accurate measure of a student's knowledge. Of course, the most essential element in the proper grading of examination papers is the teacher's serious effort to carry out his responsibiliity conscientiously, but many teachers have found a few simple suggestions about grading techniques helpful. First, a teacher should grade papers without knowing whose paper he is grading. An answer from a student who does generally good work is apt to seem more impressive than the same answer from a student who does generally poor work. Next, it is best not to grade a paper by reading it from start to finish but to read and grade all students' answers to one question at a time. This procedure ensures that a teacher will pay attention to each answer a student gives and not skim the paper after reading only the first one or two answers carefully. Furthermore, correcting papers in this way makes much less likely the possibility a teacher will alter his standards as he moves from one paper to another, for it is far easier to stabilize standards for answers to the same question than for entire examination papers. Finally, before grading a question, a teacher should list for himself the major points he expects students to mention in their answers. He can then check each essay against this list, providing yet another safeguard against altering standards as he moves from one paper to another. And such a list also provides a teacher with the means to justify his grades, since he is in a position to indicate to students what a good answer should be. Such information makes clear that grades have not been meted out arbitrarily and also aids each student in achieving both a better understanding of the material tested and an increased awareness of his own strengths and weaknesses. Of course, in order for such information to be most useful, examinations should be graded, returned to students, and discussed in class as soon as possible.

Examinations that adhere to these guidelines and avoid these pitfalls are almost sure to be reasonably successful. It should be kept in mind, however, that good examinations reinforce one another, since each one a student takes guides him in future study. Thus if he takes a number of good examinations in a single course, as that course proceeds he learns how to derive the greatest possible benefit from his study time. Multiple examinations in a single course also serve to discourage students from the popular but disastrous policy of wasting almost the entire term and then cramming for one final examination. The more frequent the examinations, the less need for cramming. Thus it is not, as some have said, that examinations encourage cramming. Infrequent examinations encourage cramming. Frequent examinations encourage studying. And good examinations encourage useful studying.

Having now discussed suitable examination procedures, we should

next consider suitable grading procedures. Much discussion has taken place about alternative grading systems, but the basic principle for constructing an effective grading system remains quite simple: it should contain the maximum number of grade levels teachers can use consistently. A grading system should be as specific as possible because grades serve as a guide for the educational decisions of both students and faculty: up to a reasonable point the more detailed the guide is, the more helpful it is. If a student's academic record is sketchy and vague, then most likely he will have a sketchy, vague idea of his own abilities and accomplishments and will be hindered in his attempts to assess his past efforts, evaluate his present capabilities, and formulate his future plans. And not only will he himself be hindered, but those who try to advise him or evaluate his accomplishments will be at a serious disadvantage. It is just not sufficient to know that Kubersky passed a course. Was he an A student, a strong B student, a weak C student, or a D student? Without an answer to this question, neither Kubersky nor anyone else knows much about his level of achievement.

But there is a limit to how specific a grading system should be. Ultimately we reach a point where no reasonable basis exists for deciding whether a student's work is at one level or another. There is little sense, for example, in trying to decide whether an English composition should receive a grade of 86.32 or 86.31, for no teacher can consistently differentiate between work on these two levels.

The question is then, using the principle that a grading system should contain the maximum number of grade levels teachers can use consistently, how many such grade levels should there be? My own experience has led me to believe that in college the most effective grading system is the traditional one, consisting of ten symbols: A, A-, B+, B, B-, C+, C, C-, D, F. This ten-level system is specific enough to provide the needed information about a student's level of achievement while enabling teachers to differentiate consistently between work on any two of the ten levels. Of course, borderline cases will sometimes arise, but the distinction between work on any two levels is clear, despite the possibility of borderline cases, just as the distinction between bald men and hirsute men is clear, despite the possibility of borderline cases.

Perhaps the most controversial aspect of the traditional ten-level system is its grade of F, for many have claimed that if a student

knows he will have a failure permanently on his record, he may become so discouraged he will give up on his education altogether. In order to preclude such a possibility it has been proposed that the grade of F be replaced by a grade of NC (No Credit), which would indicate to the registrar both that the student should receive no credit for the course and that his transcript should show no record of his having taken the course.

Such a grade, however, would obviously be pure deception, for the student did take the course and he failed to master any significant part of it. If he should take the same course again and pass it, his transcript should indicate as much. Otherwise, those who are trying to evaluate his work will be mislead, since, for example, it is likely a student who had to take introductory chemistry two, three, or four times before passing lacks the scientific or study skills of someone who passed the course on his first try. It is not a tragedy to fail a course, but it is a failure, and we must learn from failures, not give them another name and pretend they did not occur. Indeed, one mark of a mature individual is facing up to and taking responsibility for failures. As a colleague of mine once remarked during a faculty meeting in which the NC grade was being discussed: "When I die and stand before the Heavenly Judge in order to have my life evaluated, it may be that I will receive a grade of F. But let it not be said that my life was a 'No Credit'."

A suitable grading system, however, does not ensure suitable grading, for unless the system is used properly, grades will not achieve the worthwhile purposes they are intended to serve. And, unfortunately, improper uses of the system are all too common.

One such misuse is to award grades on bases other than a student's level of achievement in the course work. Irrelevant bases for grades include a student's sex, race, religion, nationality, physical appearance, dress, personality, attitudes, innate capacities, and previous academic record. None of these factors should even be considered in awarding grades. To repeat what was said earlier, a grade ought not to be a measure of a person; it ought to be a measure of a person's level of achievement in a particular course of study, and the only reasonable basis for measuring this is the quality of work which he does in that course.

The most effective way for a teacher to assure his students that no extraneous factors will enter into the awarding of grades is to state

clearly at the outset of the term exactly how final grades will be determined. How much will the final examination count? How much will short quizzes count? How about the term paper and other shorter papers? Will laboratory work count? Will a student's participation in class discussion be a factor? By answering these questions at the very beginning of the course, a teacher sets a student's mind at ease and, in addition, enables him to concentrate his time and effort on the most important aspects of the course. Of course, some teachers assume that if they do not discuss their grading policy, the students will not worry about grades. But quite to the contrary, a teacher's failure to discuss his grading policy increases uncertainty and worry and furthermore provides no guidance as to how the students should work to do their best and get the most out of the course. And, after all, such guidance is precisely what the teacher is expected to provide.

A second obvious misuse of the grading system, exceedingly rare nowadays, results from the reluctance of some teachers to award high grades. Such teachers pride themselves on how rarely they give an A or B, and how frequently they give C's, D's, or F's. But low grading is a foolish source of pride, for such grading suggests the teacher is unable to recognize good work when he sees it. That a student's work does not deserve immortal fame is no reason it does not deserve an A. Just as a third-grade student who receives an A in writing need not be the literary equal of a college student who receives an A in English composition, so a college student who receives an A in English composition, need not be the literary equal of Jonathan Swift or Bertrand Russell. Giving a student an A in a course does not mean he has learned everything there is to know about course material or that he is as knowledgeable as his teacher; giving a student an A simply means that, considering what could reasonably be expected, the student has done excellent work. If a third-grade teacher rarely gives an A or a B, his principal does not assume this teacher always has poor students in his class. He assumes, rather, that this teacher has a distorted sense of academic values. A similar conclusion should be reached about a college teacher who rarely gives an A or a B. Such a teacher is misapplying the grading symbols and preventing grades from fulfilling their important educational functions.

A third misuse of the grading system, one especially prevalent today, results from the reluctance of many teachers to award low grades. These instructors pride themselves on never giving students

a hard time or underestimating the value of a student's efforts. But high grading, like low grading, is a foolish source of pride; it suggests that the teacher is unable to recognize poor work when he sees it. Not to differentiate between two students, one doing poor or unsatisfactory work and one doing fair work, is a subtle form of discrimination against the better student. Giving a student a D or an F in a course does not mean that the student is a foolish or evil person; the low grade simply means that, considering what could reasonably be expected, the student has done poor or unsatisfactory work. If a thirdgrade teacher rarely gives low grades, his principal does not assume this teacher has the school's most brilliant students. The principal assumes, rather, that this teacher is giving his seal of approval to incompetent work. A similar conclusion should be reached about a college teacher who rarely gives low grades. Such a teacher, like the teacher who rarely gives high grades, is misapplying the grading symbols and preventing grades from fulfilling their functions.

A fourth and final misuse of the grading system is the practice commonly referred to as "grading on a curve." The essence of this widely adopted practice is deciding what percentage of students in a class will receive a particular grade, without considering the level of work actually done by any of the students. For example, a teacher may decide before a course ever begins that 10 percent of the students will receive an A, 20 percent a B, 40 percent a C, 20 percent a D, and 10 percent an F. Distributing grades in this way produces an aesthetically pleasing curve on a graph, but the procedure is invalid, for how well a student has learned a particular subject matter does not depend upon how well his fellow students have learned the same subject matter. Perhaps in many large classes approximately 10 percent of the students actually do A work and a similar percentage F work, but this fact is no reason at all why in any specific class exactly 10 percent of the students must receive an A and another 10 percent must receive an F. Suppose 25 percent of the students in a class do excellent work and 5 percent unsatisfactory work; then the 25 percent should receive an A and the 5 percent an F. Or suppose 5 percent of the students in a class do excellent work and 25 percent do unsatisfactory work; then the 5 percent should receive an A and the 25 percent should receive an F. For the grade a student receives is not to be a measure of his rank in class; it is to be a measure of his level of achievement in a particular course of study. And though judging a student's level of achievement does depend upon considering what can reasonably be expected of him, such a judgment does not and should not depend upon the level of achievement of other students who happen to be taking the same course simultaneously. Since the Procrustean practice of grading on a curve rests upon such irrelevant considerations, the practice ought to be abandoned.

Having now provided an answer to the question, "what specifically are suitable examination and grading procedures?", only one question remains for consideration: how can it be ensured that teachers will be cognizant of suitable examination and grading procedures and apply them conscientiously? The answer to the first part of this question is for those who administer graduate school programs to provide courses in methods of teaching for students intending to enter the teaching profession. These courses should be required of all students who are to be recommended for teaching positions and should include a detailed discussion of suitable examination and grading procedures. The person chosen to teach such a course ought to be himself a productive scholar and an outstanding teacher, for he is in the best possible position to make clear to graduate students that good scholarship and good teaching are not incompatible, that publishing develops a teacher's ability to think critically by leading him to submit his ideas to the judgment of his peers, while teaching encourages a scholar to express his views clearly enough to communicate them effectively to those not as knowledgeable as he.

But even if a teacher is cognizant of appropriate procedures, how can it be ensured he will apply them conscientiously? There is, of course, no practical way to ensure that anyone whether doctor, journalist, or taxi driver, will do his job conscientiously. A chairman has the responsibility to make certain no member of his department is guilty of gross negligence. But, ultimately, a teacher must decide for himself whether to be conscientious. If he is deeply committed to maintaining high academic standards, he will be willing to spend the time and effort required to make the most effective possible use of examinations and grades. But if he is unconcerned about promoting excellence and is satisfied with exalting mediocrity, he will be unwilling to give of himself in order to provide his students with effective examinations and accurate grades. What no teacher must be allowed to forget, however, is that if he chooses to ignore proper examination and grading procedures, both his students and his society will be the losers.

NOTES

- 1. Grade designations, however, are few in number and have a relattively standardized meaning. Therefore, teachers who use them idiosyncratically are not the victims of linguistic ambiguity but of pedagogic inadequacy.
- 2. Alfred North Whitehead, The Aims of Education and Other Essays (1929; reprint ed., New York: Free Press, 1967), p. 34.
- 3. These five levels of work are commonly symbolized by the letters: A, B, C, D, F. Teachers who misuse these symbols are an educational menace; their sins are discussed later in the essay.
- 4. Gilbert Highet, *The Art of Teaching* (1950; reprint ed., New York: Random House, 1954), p. 132.