

8

GRADING

And to the C students, I say you, too, can become president.
—George W. Bush, Yale University commencement

There is little to no national grading standardization in colleges and universities. Although there is universal agreement that an A grade is probably better than a C, one college's A may be a mark of rare academic distinction, while another's may simply indicate that a student has a pleasing personality and turned in her work on time. A Google search for "college grading standards" revealed every variety of policy. For example, Tacoma Community College's English department lists the detailed and specific criteria that student work must demonstrate to receive a particular letter grade. The web site of the Undergraduate Study Committee of Berkeley's Department of Electrical Engineering and Computer Science doesn't address the question of criteria to be used to award grades, but suggests a "typical" class average for a lower division course might be 2.7 (out of 4). This average might be achieved by a suggested distribution of 17% As, 50% Bs, 20% Cs, 10% Ds, and 3% Fs. The same site also notes, "In some courses, the assigned grades show wide fluctuation from section to section and/or semester, depending on who teaches the section. Furthermore, over the last 10 years, the lower and upper division course averages have inflated considerably." My university's grading definitions in its faculty manual declined in specificity over the years:

1980	2001
A: Superior	A: Superior
B: Good, above average but not superior	B: Good
C: Average	C: Fair
D: Low grade, below average, passing	D: Low grade, passing
F: Failure	F: Failure

While in 1980 an average score would earn a student a C, in 2001 an average score could be subjectively interpreted by an instructor to be “good,” thus yielding a B, and who could say nay? The seeds for “grade inflation” were thereby planted.

A few colleges issue a percentage equivalent definition for grades. For example, 90%–100% means A, 80%–89% means B, and so forth. This appears to be both a tool for standardization and an accurate representation of performance, but left unasked is the question, “Ninety percent of *what?*” Professor Jones might hold her students responsible for, say, 200 units of information. Ninety percent of that is 180 units of information. Professor Smith might hold her students responsible for 300 units of information. Eighty percent of that is 240 units of information. A performance that would produce an off-the-charts A in Professor A’s class would yield a squeaker B in Professor B’s class, yet to the outside world, both professors would seem to be using the same grading criteria. This hidden uncertainty allows a college or instructor to claim that students are being held to a high standard (i.e., “You need to have a 75% to pass this course!”) while actually allowing the student to sail through with a minimal demonstration of knowledge. The opposite, of course, can also be true—a course can appear to be deceptively easy if a low percentage score is required for a given letter grade but the students are expected to master huge amounts of information.

A handful of colleges have just pass-fail grading, and the abovementioned few have instructors report grades as a percentage, but most colleges have a categorized letter grade (A, B, C, D, F) system. Some systems are more finely divided (A, A–, B+, etc.) than others. However, the intrinsic difficulty with any categorization system is that it attempts to take what is essentially a continuum, student performance, and divide it into arbitrary segments where there is no real “natural” division between categories. Because the placement of students into these categories is so important to students’ lives, there is a tendency of colleges to pretend that the divisions between grade categories are not as arbitrary as they actually are. Scholarships, admission to graduate or professional school, and jobs may all depend on a couple of tenths of a grade point difference between students. Colleges do not like to say, “A student with a 2.75 average *may* be a better scholar than one with a 2.53, but luck has a lot

to do with it—the 2.53 student may just have had the misfortune of ending up in a lot of classes with professors who had a different grading system than did the student with the 2.75.”

If you are new to the grading process, a faculty handbook or university manual might be a place to start. Before you develop your own system for assigning students to categories, it is not a bad idea to see if your college provides any guidelines. Your institution may deal with the issue in more concrete fashion than mine. Conversation with colleagues, if they are willing to talk about it, is not a bad idea, either. If you are new to a long-established big course, it is very important to know what the grading practices have been in the past because students will tend to expect them to be the same when you take over.

As you’ve seen, a percentage system (A = 90–100, B = 80–89, etc.) gives a deceptive appearance of precision. A system that is less precise on the face of it, but more accurate in practice, relies on a series of descriptive qualifications that you can use to determine in which grade box a student should be placed. For example, a professor might define his or her grade categories as follows:

1. To earn a D (pass) in this course, a student must demonstrate a satisfactory knowledge of the vocabulary of the subject and show that he or she has, at a minimum, performed all the reading assignments. He or she should not have more than 10% unexcused absences.
2. To earn a C in this course, in addition to the above, a student will demonstrate that he or she can solve simple problems in the subject and has a basic grasp of the general principles covered in the lecture and reading assignments. All assignments shall be performed on time.
3. To earn a B in this course, in addition to the above, the student will consistently solve more difficult problems and show, by his or her questions and participation in class discussion, that he or she is fully familiar with the principles covered in assignments. The student’s writing ability, as shown in term papers, will be good and without major flaws in grammar, punctuation, and syntax.

4. To earn an A in this course, in addition to the above, the student will be able to solve difficult and complex problems. His or her written assignments will be imaginative and well thought out and he or she will demonstrate a high degree of creativity and participation in the laboratory or recitation section.

With this description of the qualities needed to earn a particular grade, assignments can then be devised to make sure that students have an opportunity to demonstrate these qualities. Obviously, if grading is based 100% on multiple-choice examinations that require memorization of facts as the only skill to be demonstrated, the category descriptions are meaningless.

Many of the words or phrases used in the descriptions above are subjective—for example, “major flaws,” “basic grasp of general principles.” Juries are faced with the same subjectivity when they are asked to determine guilt “to a certainty beyond reasonable doubt.” What is “reasonable”?

Whenever there are categories that do not have natural, clearly defined boundaries, some injustice is bound to occur. A suspect guilty in the eyes of one jury might be found innocent by another, even when the juries are presented with the same evidence. Similarly, identical student performances can be legitimately assigned different grades by different instructors.

As an individual instructor, there is little you can do about this variation between instructors. Lacking tightly defined institutional standards for performance, one can only hope that a superior student will end up with a higher grade point than an average student, but Lady Luck has a hand in this determination. What you *can* do in your own class is make sure that your grading system is understandable, equitable, and internally consistent. For large classes, manageability is also an important consideration.

Categories of Categorization

There are only two major kinds of category grading systems used in colleges. The first is the *absolute* (or “criterion referenced,” in educationese) system, where the teacher assigns students to the various grade

categories on the basis of their absolute performance. Thus, at the beginning of the semester, the teacher might say, “To earn an A in this course, by the end of the semester you must have a total percentage score that falls between 90% and 100%; to earn a B, your score must fall between 80% and 89%” and so on. Tradition and custom suggest the round numbers as the dividing points, but there is no real reason why a teacher couldn’t make it 92%–100%, or 88%–100%.

There are two characteristics of this system that you need to be aware of. First, it is inherently noncompetitive; a student’s final grade depends on his or her performance alone, not on relative position in the class. Cooperative behavior is easier to encourage in this system because helping a fellow student cannot jeopardize a student’s own grade. Second, unless the instructor is fairly familiar with the capabilities of a class in relation to the difficulties of the assignments, it is *possible* in such a system to end up with no students receiving good grades (or more students with good grades than most people would consider reasonable). Finally, students have a known and fixed mark to shoot for—get that 90% and you have an A, no matter what the class average.

In this system, the fraction of the class that falls into each grade category is determined by an interaction between 1) the ability of the class, and 2) the difficulty of the assignments and evaluation instruments. Without some experience as to the match between students and difficulty, it is possible for an instructor’s grade distributions to swing wildly from semester to semester until he or she develops a “feel” for the level of difficulty that will produce a desired distribution of grades.

If an instructor wants to use an absolute system because of the advantages it confers, yet still wishes to have a predictable and consistent fraction of the class fall in each category, he or she can do it by manipulating the cut lines between categories. There is nothing inherently sacred about the round numbers as cuts: 87%–100% for an A is as defensible as 90%–100%. Pushing the lines too far will raise eyebrows, however. If in your class the A bracket is 75%–100%, your explanations of the hellacious difficulty of your examinations may fall on deaf collegial ears.

If in your class you’d like to award the top 10% of the class As, the next 20% Bs, and so on yet still use an absolute standard, it is possible to do so by looking over several semesters’ worth of records for your course.

If the characteristics of the class haven't changed over that time, and you haven't changed the difficulty of the assignments, you should be able to see a rough pattern emerging such that if you divided A from B at 88% and B from C at 77%, you would end up with your 10%–20% distribution. Make those your absolute dividing lines in subsequent semesters.

The other major grading system is the *relative* (or “norm-referenced”) system. With this method, the instructor states, in advance, what fraction of the class will receive a certain grade, typically 10–20–40–20–10 (this assumes, of course, a normal distribution of scores). The great advantage of this system is that with no experience with the class at all, and regardless of the difficulty or ease of the assignments, the desired distribution of letter grades will be produced.

A disadvantage of the system is that with a poorly prepared class and a demanding teacher, it is entirely possible for a student to receive an A and only have a final percentage score of, say, 35%. A credibility problem is then created both for the student and instructor. How can a student have an A– (superior) grade, while knowing only 35% of the material? Conversely, with an extremely lenient instructor, a student might end up with a 95% and receive a C because half the class had better scores.

An advantage of a relative system is that if an outsider, say a potential employer, is trying to evaluate a student's performance and knows that a relative system was used, the outsider knows that a student who received an A is at the very top of the class. True, the student may be the best of a bad lot, but at least the student is at the top. In an absolute system with an easy instructor, 40% of the class might receive As, and there is no way of telling whether a student was at the top or in the middle of the class. Unfortunately for employers, professional schools, and the like, college transcripts rarely indicate the type of grading system used in a class, so this remains a theoretical advantage for those who must interpret a candidate's transcript. Some colleges are beginning to address this issue by posting class standings along with grades.

In large, lower division classes, it has traditionally been assumed that whatever the system, there will be a normal distribution of performances, and therefore grade distributions around C will be symmetrical—there will be about as many As as Fs. For a variety of reasons, this is no longer a safe assumption. Admission policies may cause a skewed or even

bimodal distribution of performances. If performance distributions are skewed to the left—that is, toward the lower performing side of the class—many grading systems will produce an F category that has two to three times more students than the A category. From a statistical standpoint, such a distribution simply reflects reality, but in an era where student retention is important to an institution's budget, such a distribution is of concern not only to faculty but also to the administration.

Resolving this difficulty for an instructor is often more a matter of being attuned to campus politics than pedagogy. As you have seen, grades in a class are whatever an instructor chooses to make them. For some instructors an A is a mark of rare distinction. Others are more generous. If *your* grade distribution is markedly different from that of the colleagues in your department who teach similar courses at the same year level, you can be assured that there will be some administrative interest in your policies. The concept of academic freedom will prohibit administrative *intrusion* into your grading practices in most cases, but there are subtle ways of conveying disapproval. Students will rarely, if ever, complain if your grading is perceived to be too easy, but if it is significantly less generous than your counterparts' policies at a different institution, whose grads your students will have to compete with for grad school admissions, you may end up sabotaging your own students' chances. Lack of national grading standards and competition for jobs and professional school admittances have led to a marked inflation in college grades.

Close examination of the absolute and relative systems will show that with suitable manipulation both can produce essentially the same results. A relative system can produce a predictable cut line between letter grades, while preserving a predetermined number of students in each category, by manipulation of the difficulty of the assignments. An absolute system can produce a predictable number of students in each category by prior manipulation of the cut lines. In both cases, some experience with the class is necessary. For instructors new to teaching large classes, I usually recommend starting with a relative system to avoid disastrous grade distributions, then switching over to an absolute system because of its non-competitive, predictable nature for students.

There are a few other grading systems that one can find in use. Instructors with class sizes on the small end of the large-class spectrum

will sometimes draw a histogram of the students' total scores at the end of the semester, then look for "natural" breaks between clusters of grades to place the divisions between letter grades. These breaks tend not to be "natural" in the sense that they represent groupings of students whose abilities are more similar to one another's than they are to students' abilities in the other groups. Rather, they are simply the statistical clumping that is the result of a relatively small class size. The only real advantage of this system is that it tends to reduce complaints from students who missed the next higher grade by one point. The disadvantage lies mostly on the student side—there is no concrete performance target for a student to aim for in his or her quest for a specific grade.

Some instructors who are comfortable with statistics use grading systems based on statistical principles. The most common starts with the assumption that student scores are normally distributed (the bell-shaped curve). The class mean, or average, is then calculated, then the standard deviation, which is a measure of the scatter of values around the mean. Some multiple of the standard deviation is then used to set the boundaries of the grades. For example, if a multiple of 1.3 times the standard deviation is used to set the lower boundary of A grades, and 1.0 times the standard deviation to set each lower grade boundary, the result will be 7% As, 24% Bs, 38% Cs, 24% Ds, and 7% Fs. If there is *not* a normal distribution, for example if it is skewed or bimodal, then this method loses the mathematical underpinnings of its use.

Such a system has the advantage of looking "scientific" and hence generates few student complaints. However, at its heart, it is just as arbitrary as any other method. Why not use 1.2, or .9, as the multiple to set the boundaries in order to get the grade distribution you feel comfortable with? No reason at all.

Non-Normal Distributions

Large classes *should* display a normal distribution of performance, with relatively small (and equal) numbers of students at the extremes, with the bulk of the class centered around the class average. However, it is becom-

ing increasingly common at the kinds of colleges that have large classes to see a flattened distribution, with more students at the extremes than one would expect, or skewed (typically toward the low end) or even bimodal distributions with peaks at the extremes, and a flat spot in the center. These situations can produce great difficulties for instructors.

With a class that has a normal distribution, you can't go far astray by tailoring the level of your presentation toward the middle of the class. The greatest number of students are in this category, and the relatively small number of extremes aren't that far from the middle. However, with the worst-case bimodal distribution, there are very few students in the middle, and if you aim at either of the extremes, the other extreme is going to be either terminally bored or impossibly confused.

I'm not aware of any truly satisfactory ways of dealing with this. When it has happened in my big classes, I've tried a few techniques that have had some success, but all of them require an enormous amount of time.

To give me early warning of what the situation is going to be, I schedule a "mini-exam," only two and a half weeks into the semester. This is really just a quiz, but it has the trappings of an examination—same format, same type of questions, and so forth. It is not for "practice": It counts toward the grade, but not much, so if a student blows it, no real harm is done. The only difference between it and a regular examination is that it has only 10 questions and is based on a couple of chapters. Based on the experience of a couple of years, I can then make some predictions about what the distribution of scores will be on the first regular examination and prepare the difficulty level accordingly.

After the first full examination, those students who fail are told that they *must* come in to see me for a personal office visit and study consultation. The students are told to come prepared to answer three questions: 1) Why do *you* (the student) think you failed the examination? 2) What specific steps do *you* propose to take to make sure it doesn't happen again? 3) What can *I* (the instructor) do to help you? This discussion is very illuminating because students often are clueless about the reasons for their failure—the same amount and kind of study they are currently doing brought them As in high school. I spend 5–10 minutes with each one. In my larger classes, I might have to see more than 125 students in a two-week period, leaving little time for anything else. Does it help? In the first

semester after I started doing it, my end-of-semester failure rate dropped by one-third and that number has somewhat improved in subsequent semesters—all without lowering grading standards. Is this an appropriate use of a senior professor's time at a research university? As I have pointed out to my administrator colleagues, if this procedure keeps just five students in school, who would otherwise have flunked out, dropped out, or transferred, I have saved the university in that one semester more money than I would have generated in research overhead for the whole year in my field. Although the students pick up some study tips, I have the feeling that the real reason the technique works is that the students have the feeling that even though it is a big class, someone actually *cares* whether they do well.

For the better students in class, I have a standing offer that if they're bored with the regular assignments and can demonstrate to me that they already have a mastery of the subject, I'll work with them on a custom package of readings, consultations, lab work, or whatever they want so that they can learn something new—all without a grade penalty. I have the occasional student take me up on it, but what I hear from these students when I run into them again in upper division courses is that they would rather take an easy, no-effort A and spend the liberated time on their more challenging courses or outside activities. As a result, I have a fairly easy conscience about neglecting the top group and spending disproportionately more time with the ones hanging on by their fingernails.

Special Considerations

Makeups

There are always students who miss assignments and examinations, sometimes for good cause, and you will need to establish a consistent policy about makeup work before the semester begins. Some instructors don't have a makeup policy—if a student misses an exam, that's his or her problem. If you are going to permit makeup work, however, there are several alternatives. If you have a number of roughly equal-weight assignments, say three one-hour exams or weekly quizzes, you can allow students to automatically drop their lowest score and the percentage will

be calculated on the revised total (you need to realize that this artificially raises the class average and may have an impact on your distribution of letter grades). If a student misses an exam, that becomes his or her de facto lowest score and is dropped automatically—you don't need to bother with excuses or doctor's notes. However, what do you do if a student misses a *second* assignment?

You can always do nothing. Into each life a little rain must fall. If you are sympathetic to the student's case, you don't want to drop a second exam and calculate a percentage on the balance of available points because that might actually give the student an *advantage* by missing the assignment. On the other hand, if the student has satisfactorily completed a fair number of other, similar assignments, you can prorate the missed assignment—calculate the average score, say 75%, on completed assignments, and figure that the student *would have done* about as well on the missed assignment as on the completed ones. You therefore give the student the average percentage on the missed assignment. If it were a missed 10-point quiz, you'd give the student 7.5 points. If the missed assignment is unique, say a term paper or a single mid-semester examination, then it is difficult to impossible to apply this technique.

I go about handling makeups in a slightly different way. If students miss an exam, I simply let them take a makeup examination—but it is always in essay format. Clearly, they can't be given the same multiple-choice examination that all the other students took because the answers are all over the campus five minutes after the examination is over. At the same time, I don't want to make up a 40-question multiple-choice examination for just one student. Instead, all the student has to do is make an appointment with me anytime between the end of the missed examination and the date of the next examination and take the makeup at a time of mutual convenience. I don't have to be a cop checking on the authenticity of their excuse with this method, although obviously if the student has spent an enormous amount of time developing a creative excuse to con me into allowing him or her to take a makeup, I would be a cad indeed if I didn't listen to it, stroke my chin, sympathize about the tragic loss of their cousin in the awful accident, and then allow them to take the makeup. In my case, once the word got around that it was easy to get a makeup, but the makeup was an essay, there

was an amazing drop in the number of grandmothers who unexpectedly passed away during the semester.

Borderlines

If your class is big enough, there will be substantial numbers of students who are one point away from the next higher grade by the end of the semester. You may, with perfect justification, express to your students at the beginning of the semester the idea that such is the nature of life, close doesn't count, 89.9999 is a B+, and that's the way it is.

On the other hand, especially in first-year courses, it is quite common for students to start badly because they have “freshmanitis,” get with the program, and end up doing quite well. However, because of their bad early performance, an A performance on the final might be pulled down to a B– for the semester, once the earlier scores are factored in. You may wish to recognize this improvement in an individual student by departing slightly from the fixed relationship between numerical score (or percentage) and letter grade. Unless you handle this carefully you will open yourself up to an endless stream of students who heard about their lucky compatriots and will beseech you with plea bargaining requests after the semester is over.

I handle this by incorporating into the printed course rules two rare exceptions to the usual rule that percentage score equals letter grade. The first exception is where I have personal knowledge that a student who had received otherwise good scores had blown an examination through circumstances beyond his or her control, for example having gotten news of a family death just before an examination. If the throw-out-the-worst-exam system were used, this exception would not be necessary. It would seem that students would abuse this bit of largesse, but I have not found it to be the case in practice.

The second exception is where a student is a point below the next higher grade, and the final examination is “substantially” better than his or her average scores. In this case, I'll consider a boost to the next higher grade. *Substantial* is a subjective term, of course, and allows me a little executive leeway.

In a class of 300, perhaps a half-dozen students will meet the qualifications. Adopting this system means that two students could have iden-

tical final scores but different final letter grades, but I've not had any squawks from students about this—they seem to understand that the system recognizes improvement and thus can only be to their advantage.

Multiple-Section Grading

There are two cases where several parties might be involved in the grading process. In the first, there is a lecturer who supervises a number of graduate students teaching recitation or laboratory sections who feed recitation or lab scores to the course instructor. In the second, there might be several lecturers teaching different lecture sections of the same course simultaneously, or several instructors sequentially lecturing to the same group of students, a situation sometimes called “team teaching,” although there may be precious little teamwork involved.

The first case presents few problems besides administrative and managerial difficulties. Grades can be normalized between sections as described in Chapter 4, and the assistants are really assistants, whose job is to help you. The administrative structure is basically hierarchical, and you are the Boss.

When multiple faculty are involved in a single course, the situation is more delicate. Faculty members tend to have strong and diverse opinions about testing and grading, but only one grade per student can be turned in at the end of the semester.

Like-minded faculty might get together and decide, “Let's do a course on XY! I can do the X lectures, you do the Ys, we'll both sit in on each other's lectures, and we can write the exams together. After all these years talking with you, we have similar ideas on how students should be graded, and it'll be a ball!” And it would be. However, a more typical multiple-instructor arrangement is something cobbled together by a department chair who has instructors who need bits and pieces of teaching credits to make up their teaching load. The participants may loathe each other after years of departmental squabbling and have diametrically opposite views on teaching, testing, and grading.

Differences in lecture styles present little difficulty—students always have mixed reactions to any one style, and a student who strongly dislikes the first lecturer in a series may love the second, while his or her seatmate has just the opposite reaction. The stickiness arises not from the lecture,

but in grading.

Some departments may have multiple lecturers, but all teach from a standardized syllabus, and their students take common examinations prepared as a result of negotiations between the lecturers. Such an arrangement can have a stifling effect on lecturer creativity but eliminates differences in grade performance between sections caused by differences of opinion between the participants about grading practice. However, it also tends to increase the practice of “teaching to the test.” Another unwelcome side effect is that if an instructor finds that his or her lecture section is having a particularly hard time dealing with a concept, it is difficult for that instructor to devote extra time to the subject without having the class fall behind the syllabus and possibly be at risk on the next examination.

Where there are multiple lecturers, each of whom makes up his or her own lecture examinations, some of these problems are addressed, but it is unlikely that randomly selected lecturers will have the same grading theory, so the distribution of grades between the sections can differ, sometimes widely.

Students readily compare notes, and it is likely that the “tougher” instructor will draw more student complaints about grading than if the policies were the same, but he or she was the sole lecturer. What happens after that depends on too many factors to be able to make a general prediction. First-year students are more affected by their perceptions of an instructor’s grading practices than advanced students, and if students have a choice between an “easy” instructor and a “hard” one for the same course, all other things being equal there will be a tendency to gravitate toward the easier one’s sections. If tenure and promotion decisions at an institution are strongly influenced by student evaluations, an untenured instructor who finds that his or her standards are markedly higher than a co-teacher in a course may have a strong personal disincentive to continue with that practice.

Squawks

No matter how well thought out your grading procedure, there will be complaints about it. That’s just human nature. Students are unlikely to come to you and say, “You gave me a B+ and I only deserved a C+. Would you change it, please?” The reverse request can be expected. By their

nature, complaints about grades must be handled individually, but you will want to remember that resolving complaints in big courses is very much like case law in the legal world—once a precedent has been established, it is very difficult to deny the next plaintiff equal relief. The student communication network is unbelievably efficient, and once it is known that a certain approach has been successful with you there will be others to follow.

Unless you *like* the idea of dealing with a parade of students outside your door when the final grades are posted, your best grading policy is one that is simple, clear, unambiguous, and delivered in writing at the beginning of the semester. This policy should not be so rigid that you cannot accommodate the occasional legitimate squawk, but should be tight enough to prevent endless lines of supplicants in the future.

Many colleges have formal procedures in place for students to object to what they consider “unfair” grades. Typically a student will go to the professor first, then the department chair, then the dean and/or some kind of campus judicial board set up to handle grading disputes, and finally a campus ombudsman, if there is one. In recent years, parents often run interference for their kids, adding a new and delicate dimension to the issue. When I was department chair, I used to hate to get involved in these matters. More often than not, the student would have a legitimate grouse because the grading instructions in the course were vague, contradictory, or not administered in an evenhanded way. “Academic freedom,” however, said that I couldn’t butt in (nor could the dean) if the instructor was intransigent, and it would be up to me to explain to the student that, yes, they did have a point, but no, there wasn’t anything I could do about it. Needless to say, my feelings toward the faculty member who had put me in this position were not warm.

According to our ombudsman, the vast majority of disputed grade cases that make it to the office involve not judgment calls about, say, how a paper should have been graded, but unclear grading policies that allowed some degree of interpretation on the student’s part. Naturally, when given this opportunity, students will interpret things in their favor and be unpleasantly surprised when, after the fact, the instructor doesn’t allow that view.

Students rarely win disputed grade cases, but they can involve an outrageous amount of faculty time (and emotional stress), and as a large-class instructor, you are far more likely to face them than you would in a small class. Again, the very best way to avoid them is to have a clear, specific, written grading policy that is issued to the students on the first day of class and allows very little wiggle room in its interpretation.

Students will remember the grade you issue to them long after they have forgotten most of the facts that you presented. Like it or not, grades can have a profound effect on your students' lives, so the humane, professional teacher will spend considerable attention and effort in making sure that the grading system is as equitable, valid, and reliable as possible.