Calculating and Assigning Grades

Grades may be assigned in a variety of ways, each with its own strengths and weaknesses. Which model you choose depends on your department's policies, the size and type of course you teach, and your views of education and the purpose of grades. The suggestions below are designed to increase your understanding of your options and point out the pitfalls of certain types of grading strategies. See "Multiple-Choice and Matching Tests" and "Short-Answer and Essay Tests" for information about grading those types of exams.

**General Strategies**

**Familiarize yourself with department standards.** Check to see how grading has been handled for the course in past semesters; if possible, see whether you can obtain past class grade distributions. Ask colleagues who have taught the course before about their grading criteria and general class performance.

**Relate department standards to your own conception of the course.** Identify the objectives or goals you want your students to meet. What skills and knowledge are absolutely essential for students to pass the course? What would you wish from an A student?

**Weight various course components in proportion to their importance.** Quizzes should count less than a three-hour exam, but if you make the final exam worth 60 percent of students' final grade in the course, you encourage students to cram at the end rather than work at an even pace throughout the term (Lowman, 1984). As a rule of thumb, the final should count for no more than one-third of the course grade.

**Make the institution's definitions of grades known to students.** Refer students to the school catalogue for an explanation of your institution's grading scales. Here is the scale used at the University of California at Berkeley:
Homework

**Give frequent assignments.** Asking students to turn in frequent short assignments has a number of advantages (Committee on the Teaching of Undergraduate Mathematics, 1979):

- You have continual opportunities to see how your students are doing.
- Students become accustomed to regular and systematic study and tend to procrastinate less.
- Students acquire a clear idea of what sorts of problems or assignments they should be able to do.

You need not grade every assignment. Instead, grade one or two problems without telling the students in advance which ones those will be. Or collect two or three problems a week for grading. Some faculty ask their students to accumulate homework in a notebook, which is called in for checking from time to time, or they give short quizzes on the problem sets and grade those. For assignments you do not grade, distribute an answer sheet on the day the homework is due so that students can check their work. (Source: Committee on the Teaching of Undergraduate Mathematics, 1979)

**Give students tips on how to solve problems.** According to Whitman (1983), research shows that the key difference between novice and experienced problem solvers is that experts have more immediate access to a wider repertory of skills. Novice problem solvers take longer to locate appropriate strategies and proceed in slow, step-by-step fashion. Help students figure out how to approach an unfamiliar or difficult problem. Here are some suggestions (adapted from Andrews, 1989; Brown and Atkins, 1988, p. 185):

- Write out the information specifically requested by the problem.
- List all the givens, both explicit and implicit.
- Distinguish the key points.
- Try to explain the problem to someone else.
- Make a flowchart with yes/no options, draw a diagram, or represent the problem graphically or mathematically.
- Think of similar problems you have solved successfully.
- Break the problem into smaller parts.
- Do the easiest parts or steps first.
- Make a rough approximation of what the solution should look like.
- Work backward from the goal.
- Work backward and forward from the midpoint.
- Systematically use trial and error.

---

*Calculating and Assigning Grades*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Excellent</td>
</tr>
<tr>
<td>B+</td>
<td>Good</td>
</tr>
<tr>
<td>B</td>
<td>Fair</td>
</tr>
<tr>
<td>C+</td>
<td>Poor; barely passed</td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D+</td>
<td>Passed (the equivalent of a C - or better)</td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Fail</td>
</tr>
<tr>
<td>P</td>
<td>Not passed (the equivalent of a D + or worse)</td>
</tr>
<tr>
<td>NP</td>
<td>Satisfactory (for graduate students only: passed at the</td>
</tr>
</tbody>
</table>
level of B - or better)

U Unsatisfactory (for graduate students only: the equivalent of C + or worse)

I Incomplete (work of passing quality but a small portion is incomplete, such as a lab experiment or term paper; the student must make arrangements to complete the work with the instructor before an incomplete can be issued)

IP In progress; final grade to be assigned upon completion of two-term course sequence

Use software to compute and keep track of students' grades. Some faculty keep students' grades on spreadsheets, such as Microsoft Excel, that permit sorting, statistical analysis, graphing, and other computations. Faculty who use computerized grading report that it saves considerable time and effort, and many prefer the flexibility of commercially available spreadsheets to customized grading programs. Straka (1986) evaluates the applicability of various computer spreadsheet programs for use as computerized grade books. Magnan (1989) identifies key characteristics of effective software programs (for example, able to handle missing values or temporary zeros for assignments not yet turned in or exams to be made up; able to drop the lowest grade) and offers advice on how to select software.

Approaches to Grading

Criterion-referenced grading versus norm-referenced grading. The principle underlying criterion-referenced grading is that a student's grade reflects his or her level of achievement, independent of how other students in the class have performed. If all the students in a seminar give strong oral presentations, they will all receive A's or B's. Conversely, if none of the students in a class scores better than 80 percent on a midterm exam, then no one in the class receives a grade higher than, say, B - on the exam. Under norm-referenced grading, in contrast, a student's grade reflects his or her level
of achievement relative to other students in the class. Only a certain percentage of the class will receive A's, a good portion will receive C's, and at least some will receive D's and F's. Norm-referenced models are often called grading on the curve.

Researchers conclude that when students' test scores are fairly well distributed across the range of possible points, it does not necessarily matter which model you employ (Svinicki, n.d.). However, when the overall performance of the class is low or high, it matters a great deal which model you use. When many students have done well on an exam, they would want some form of criterion-referenced grading so that everyone who did well will receive an A or B. When many students have done poorly, they would want grades assigned by curve so that at least some would receive A's or B's. Of course, other factors besides the performance of the class will affect the model you choose, including your educational philosophy and the importance of students' mastering the content being tested. In general, however, grading systems based on criterion-referenced standards are more defensible than those that rely strictly on the curve.

Grading on the basis of improvement. Some faculty believe that course grades should take into account the amount of growth and development a student shows over the course of a semester. Otherwise, these faculty argue, a student who enters the course fairly knowledgeable will receive an A even if he or she learns very little or demonstrates little effort or improvement. McKeachie (1986), Pollio and Humphreys (1988), and Terwilliger (1971) point out that grading largely on the basis of progress produces dire inequities: a student who comes into the course with the least background might still be the poorest student in the class at the end of the course but might get an A for progress; a student who shows little growth may still be outstanding and yet get a C for progress. Grading on the basis of improvement also makes it difficult for students to interpret what their grades mean: does a B mean that their work is above average or that their improvement is above average? It is also a disservice to students whose knowledge is inadequate for a higher-level course to receive a B in your course. McKeachie (1986) recommends that a student who performs poorly but has tried and has made some progress be given a D rather than an F; assign the Fs to students who demonstrate low achievement and have made little progress. Some faculty take improvement into account by giving extra points if a student scores higher on the second midterm than on the first, and by giving a modest number of bonus points at the end of the term if a student's improvement has been steady throughout the semester.
Calculating and Assigning Grades

Self-grading and peer grading. Some faculty let students grade themselves. These faculty ask students to justify in detail the reasons for the grade, taking into account their performance on exams and assignments, their perceived grasp of the material, the amount of time spent on the course, and the amount of reading completed. Though this approach does develop students' abilities to evaluate their work, it takes away from faculty one of their chief responsibilities: to make professional judgments about students' learning and tell students how well they are performing. A variation of self-grading is peer grading, where students grade one another's work. This procedure works best in classes that feature a lot of small group work that enables students to judge how well other students are performing. If you wish to try this strategy, see "Collaborative Learning" for suggestions. (Sources: Jacobs and Chase, 1992; Fuhrmann and Grasha, 1983)

Models of Grading Student Performance Relative to Others In the Class

Grading on a curve based on the class's performance. In this norm-referenced model, grades are determined by comparing a student's overall performance with that of other students. Students' scores are arrayed from highest to lowest, and the grades students receive depend on where they land in the array. The key, of course, is where to set the cutoff points. Some instructors determine beforehand which percentages of students will receive A's, B's, C's, D's, and Fs. So, for example, if you determine that 20 percent of the class will receive A's, you simply count down from the highest score until you reach the number of students that corresponds to 20 percent. The choice of percentages is left to your judgment. Some faculty use the mean and standard deviation to determine cutoff points. For example, A's and B's are assigned to students who score a specific amount above the mean, C's to students whose scores fall close to the mean, and D's and Fs to those whose scores fall a specific amount below the mean. Some researchers (for example, Gronlund, 1974) provide guidelines, such as 10 to 20 percent A's, 20 to 30 percent B's, 40 to 50 percent C's, 10 to 20 percent D's, 0 to 10 percent Fs.

This model has the advantage of rewarding students whose academic performance is outstanding in comparison to their peers, but it has several major drawbacks (Frisbie, Diamond, and Ory, 1979). The grade doesn't really indicate how much or how little students have learned—only where they stand in relationship to the class. In addition, no matter how strong the class is, some students will receive low grades; or, no matter how weak the class is, some students will receive high grades. Further, your grading standards will
fluctuate with each group of students—a student whose work earns a C+ in the fall term might have received a B- the spring term before. Some faculty try to compensate for inequities by adjusting the cutoff scores or by assigning a higher percentage of A's than usual if the class is really good. But these adjustments still arbitrarily limit the number of A's, so that no matter how much students actually learn, only so many of them will receive A's. (In contrast, other methods of grading permit all students who work hard to earn high grades.) Moreover, researchers Hanna and Cashin (1988) believe that grading on the curve seems to encourage exclusion, isolation, and competitiveness.

**Grading according to course or departmental practices or according to faculty consensus.** Some faculty try to have their grade distributions reflect the averages reported in their department. Brown (cited in Johnson, 1989) reports that faculty tend to distribute grades as follows: 22 percent A's, 35 percent B's, 29 percent C's, 10 percent D's, and 4 percent F's. Hanna and Cashin (1988) suggest that all faculty who teach the same course might develop a consensus on the distribution of grades suitable for a typical class, say, 20 percent A's, 25 percent B's, 30 percent C's, 20 percent D's, and 5 percent F's, although the instructors would be expected to deviate from the typical distribution to reflect the uniqueness of their class (higher than average performance or less well prepared).

**Grading according to breaks in the distribution.** In this model, you array the scores from highest to lowest and look for natural gaps or breaks in the distribution. For example, on a midterm six students score 80 or higher and two students score 72; no one scores between 79 and 73. Instructors using this model will assign A's to students who scored 80 and above, and start the B's at 72. The assumption here is that these breaks represent true differences in achievement. However, given the unreliability of teacher-designed tests, the breaks could occur purely by chance, or they could be caused by guessing or poorly written items. Further, the grade distribution depends on judgments made after students have taken the test rather than on preestablished guidelines that can be stated prior to testing. This model is not recommended. (Source: Jacobs and Chase, 1992)

**Grading on a bell-shaped curve.** A bell curve, or a "normal" distribution curve, is a symmetrical statistical model. A small percentage of the class receives A's and F's, a larger percentage receives B's and D's, and most students receive C's. If you were to use a pure normal curve, the distribution of grades would be 7 percent A's, 24 percent B's, 38 percent C's, 24 percent
Calculating and Assigning Grades

D's, 7 percent Fs. Though bell curves have their uses, they are wholly inappropriate for grading. Student performance is not necessarily normally distributed within a class, and teacher-made tests are almost never so well designed as to yield such distributions. (Sources: Gronlund, 1974; Terwilliger, 1971)

Models of Grading Student Performance Relative to a Standard

Grading according to absolute standards. In this criterion-referenced model, a student's performance is compared to a specified, fixed standard set by the instructor. For example, according to McKeachie (1986), if a test has 150 possible points and you want to set standards of 90 to 100 percent = A, 80 to 89 percent = B, and so on, the distribution might look like this:

- 140 and above (93 percent of 150) = A
- 135 to 139 (90 percent) = A-
- 131 to 134 (87 percent) = B+
- 125 to 130 (83 percent) = B
- 120 to 124 (80 percent) = B-

and so on

An advantage of this approach is that any number of students may earn A's and B's. You might also indicate to students that you may grade more generously than the announced standards but never tougher. A major problem with the model is how to set the standards in a rational, legitimate way. Another problem arises if many students perform very poorly (say, only one student scores above 120). At that point, an instructor could reset the standards to reflect students' performance, but such tinkering confuses the meaning of grades and may frustrate students. (Source: Terwilliger, 1971)

Grading according to highest scores earned and percentages thereof. This model (developed by Carter as reported in Fuhrmann and Grasha, 1983, p. 184, and also described in Svinicki, n.d.) is a hybrid criterion-referenced and norm-referenced approach that combines the advantages of each. You assign grades on the basis of the highest scores earned in the class. It works like this:

- Compute a score for each student.
- Compute the mean score of the best-performing portion of the class. If you have a superior class, you may want to use the scores of the upper
15 or 20 percent; if the class is less capable than previous classes, you might use the top 5 to 8 percent of the distribution. For an average class, use 10 percent. To calculate the mean, add together all the scores in the best-performing portion of the class and then divide by the number of scores in this sample. • Assign grades according to some predetermined scale; for example: A = 95 percent of the mean of the best-performing portion of the class, B = 85 percent of the mean, C = 75 percent of the mean, D = 65 percent of the mean.

In this model, class performance plays a role in determining the score needed for each grade, but the number of students who can earn each grade is not limited. Some faculty take shortcuts with this model by simply using the highest score (rather than the mean of the highest scores) so that 90 to 100 percent of the highest score is awarded an A, 80 to 89 percent a B, and so on. The drawback of this shortcut is that it is too dependent on a single student's score.

Grading according to mastery of objectives. For this criterion-referenced model, you first prepare a list of detailed objectives, the measurable skills and knowledge students are expected to attain. Students' performance is evaluated on whether or not they have mastered these objectives. The premise here is that a grade indicates how much a student knows rather than how many other students have mastered more or less of that domain. It is possible that all students in a class could receive A's and B's. There are two clear advantages of this model of grading (Frisbie, Diamond, and Ory, 1979): (1) most students who work hard enough and receive good instruction can obtain good grades; (2) the focus is on achieving course goals, not on competing with the other students. The drawbacks of this model, however, are that the instructor must be able to specify clearly the levels of knowledge and skills students must master to earn a given grade and that the instructor must be able to determine the minimum level of performance necessary to attain each grade (Frisbie, Diamond, and Ory, 1979). In many college courses, as Hanna and Cashin (1988) point out, the content is so extensive that an instructor cannot specify the requisite knowledge and skills with precision.

**Calculation of Final Grades for the Course**

**Begin with numerical scores.** Calculate final grades by converting letter grades to numerical equivalents if you need to. If you have assigned points
Calculating and Assigning Grades

throughout the term to students' papers and tests, use those scores. If you need to convert, say, a grade on a research paper, use A = 95, A- = 90, B+ = 87, and so on.

**When course tests and assignments are all equally weighted, add them all up to obtain a total score.** Many faculty calculate final grades by totaling the scores obtained by each student on each of the course requirements (for example, midterms, papers, final exams), putting these total scores in numerical order, and setting a cutoff point for each final letter grade. However, there are some technical drawbacks to this straightforward procedure (Jacobs and Chase, 1992). For example, a test with a wide spread of scores will more heavily influence the final grades than a test with a narrower spread. This problem can be overcome by converting students' raw scores into standardized scores. (See Fuhrmann and Grasha, 1983, pp. 186-188, for a formula for transforming raw scores into standardized scores. Computerized test grading and analytic programs greatly ease the calculations.) Some researchers argue against transforming raw scores, believing that little is gained by such arithmetic maneuvering, and the probability of error increases. McKeachie (1986), for example, stresses that adding raw scores is adequate for most courses.

**When course tests and assignments are not equally weighted, calculate weights for each course component and convert individual scores on course tasks.** Here's a method for calculating weights (adapted from Frye, 1989, pp. 187-189). First, write down the weights you announced to your class; for example:

- Two midterms 50 percent
- Two lab projects 30 percent
- Final exam 20 percent

Then convert these percentages to points, allowing 100 percent to equal 1,000 points. Thus, the two midterms are worth 500 points, the two lab projects 300 points, and the final exam 200 points. Since each of the midterms is worth 250 points (500 divided by 2), you need to adjust students' actual scores on the midterm to weight them for the final course grades. Suppose the first midterm had 40 items and the second midterm had 50 items:

- midterm 1: 250 divided by 40 = 6.25 points per item
- midterm 2: 250 divided by 50 = 5.00 points per item
If a student has 31 correct answers on the first midterm, she would receive \(31 \times 6.25 = 193.75\) points; for having 42 correct answers on the second midterm, she would receive \(42 \times 5 = 210\) points. Convert the lab projects and final exam in a similar fashion, and use the point totals to compute final grades.

One professor adjusts the weights assigned to various course components to reflect students' performance. For example, if a course has two midterms and a final, the student's highest score on the three tests is weighted 50 percent, the middle score 30 percent, and the lowest 20 percent. The advantage is that students know their best work counts more heavily in their final grade. The disadvantage is that students do not know in advance how much each course component will be weighted.

Set cutoff points. There are four general ways to set cutoff points: straight percentages, standard deviations, percentages of the highest ranges of scores, or absolute standards. With straight percentages, you set the grade distribution on the basis of your judgment of department norms or faculty practices. For example, suppose you decide that 15 percent of your students will receive A's, 20 percent B's, 45 percent C's, 15 percent D's, and 5 percent F's. Then you count down the total scores and assign the grades. Because this procedure limits the number of A's and B's, it is not recommended.

For the same reason, using standard deviations to calculate cutoff points is also not recommended. This approach, which requires some mathematical skill or computer software, entails calculating the mean score and the statistical standard deviation from the mean. Students whose point total is within, say, one-half a standard deviation from the mean receive C's. The cutoff between A's and B's would lie one standard deviation above the upper cutoff of the C's, and the cutoff between D's and F's would lie one standard deviation below the lower cutoff of the C's.

A better way to set the cutoff points is to use a percentage of the distribution (similar to grading according to highest scores earned and percentages thereof, described earlier). This strategy does not restrict the number of A's and B's. Here, you take the upper 10 or 20 percent of the distribution, calculate the mean, and then calculate a percentage of the mean to set the cutoff points (90 percent of the mean is an A, 80 percent a B, and so on). The width of the band and the percentages of the mean for each letter grade are up to you. In calculating grades, you may wish to take into account natural breaks in the distribution that are slightly off from your ideal percentages.
Calculating and Assigning Grades

Another way to set cutoff points is to adopt an absolute standard rather than relying on the performance of the class. The premise here is that you can specify the point totals necessary for student achievement at various grade levels, representing various levels of mastery of the material.

Consider students' progress over the term. For students on the margin, you could consider such factors as improvement over the course of the semester. One math professor quantifies students' improvement over the semester. Say a class has two midterms worth 100 points each and a final worth 200 points. Ole's score is 50 out of 100 on the first midterm, 80 out of 100 on the second midterm, and 190 out of 200 on the final. His unadjusted total score for the course is 320. To take into account Ole's steady improvement, the professor weights Ole's scores for the second midterm and final more heavily. The weight of the second midterm is calculated by subtracting Ole's score on the first midterm from the total points available for both first and second midterms, or 200 - 50 = 150. His actual score on the second midterm is then multiplied by this weighting factor. Thus his adjusted score for the second midterm is (200 - 50)(80/100) = 120. His first score and his second, adjusted score are then added together, 50 + 120 = 170. His adjusted score on the final is calculated using the same process. Ole's cumulative total score (170) is subtracted from the total points available for both midterms and the final (400). His actual score on the final is then multiplied by this new weighting factor, or (400 - 170)(190/200) = 218.5. To calculate the total adjusted score, add his adjusted score on the midterms (170) to his adjusted score on the final (218.5) for a total of 388.5.

References

TOOLS FOR TEACHING


Magnan, B. "How to Make the Grade." *Academic Leader,* 1989,5(10), 1-3, 5,10.


Straka, W. C. "Spreadsheet-Gradebook Connection." *Teaching at the University of Nebraska, Lincoln,* 1986, 7(3), 1-3,7. (Newsletter available from the Teaching and Learning Center, University of Nebraska, Lincoln)

Svinicki, M. D. *Evaluating and Grading Students.* Austin: Center for Teaching Effectiveness, University of Texas, n.d.