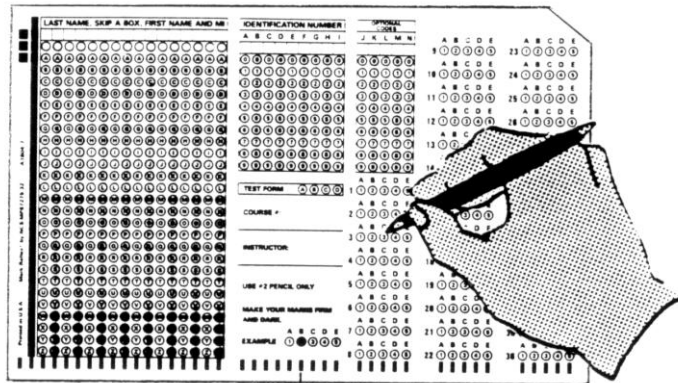


Assigning Course Grades

Technical Bulletin #5



Evaluation and Examination Service
The University of Iowa
(319) 335-0356
Revised July 1996

This paper is intended to be used as a reference by instructors who wish to establish grading policies and practices or to review and revise their current grading procedures. Much of the content of this paper appeared in a booklet of the same title prepared by the author and others at the Measurement and Research Division, University of Illinois. The viewpoints expressed herein are those of the author. This booklet has been divided into the following sections:

I.	Introduction	2
II.	Grading Comparisons	3
III.	Grading as a Part of Evaluation	6
IV.	Basic Grading Guidelines	8
V.	Methods of Assigning Course Grades	11
VI.	Grading in Multi-Sections Courses.....	18
VII.	Evaluating Grading Policies.....	20
VIII.	Assistance Offered by the Evaluation and Examination Service (EES).....	21
IX.	Appendix A - Basic Statistical Concepts.	23
X.	References for Further Reading	25

I. INTRODUCTION

The end-of-course grades assigned by instructors are intended to convey the level of achievement of each student in the class. These grades are used by students, other faculty, university administrators, and prospective employers to make a multitude of different decisions. Unless instructors use generally-accepted policies and practices in assigning grades, these grades are apt to convey misinformation and lead the decision-maker astray. When grading policies and practices are carefully formulated and reviewed periodically, they can serve well the many purposes for which they are intended.

What factors should a faculty member consider in establishing sound grading policies and practices? The issues that contribute to making grading a controversial topic are primarily philosophical in nature. There are no research studies that can answer questions like: What should an "A" grade mean? What percent of the students in my class should receive a "C"? Should spelling and grammar be judged in assigning a grade to a paper? What should a course grade represent? These "should" questions require value judgments rather than an interpretation of research data; the answer to each will vary from instructor to instructor. But all instructors must ask similar questions and find acceptable answers to them in establishing their own grading policies. It is not sufficient to have some method of assigning grades--the method employed must be defensible by the user in terms of his or her beliefs about the goals of a university education and tempered by the realities of the setting in which grades are given. An instructor's view of the role of higher education consciously or unknowingly affects grading plans. The instructor who believes that the end product of a university education should be a "prestigious" group which has survived four or more years of culling and sorting has different grading policies from the instructor who believes that most college-aged youth should be able to earn a college degree in four or more years.

An instructor's beliefs are influenced by many factors. As any of these factors change there may be a corresponding change in belief. The type of instructional strategy used in teaching dictates, to some extent, the type of grading procedures to use. For example, a mastery learning approach to teaching is incongruent with a grading approach that is based on competition for an established number of "A" or "B" grades. Grading policies of the department, college, or campus may limit the procedures that can be used and force a basic grading plan on each instructor in that administrative unit. The recent response to grade inflation has caused some faculty, individually and collectively, to alter their philosophies and procedures. Pressure from colleagues to give lower or higher grades often causes some faculty members to operate in

conflict with their own views. Student grade expectations and the need for positive student evaluations of instruction both probably contribute to the shaping or altering of the grading philosophies of some faculty. The dissonance created by institutional restraints probably contributes to the wide-spread feeling that end-of-course grading is one of the least pleasant tasks facing a college instructor.

With careful thought and periodic review, most instructors can develop satisfactory, defensible grading policies and procedures. To this end, several of the key issues associated with grading are identified in the sections that follow. In each case, alternative viewpoints are described and advantages and disadvantages noted.

II. GRADING COMPARISONS

Some kind of comparison must be made when grades are assigned. For example, an instructor may compare a student's performance to that of his or her classmates, to standards of excellence (i.e., pre-determined objectives, contracts, professional standards) or to combinations of each. Common comparisons used to determine course grades and some advantages and disadvantages of each are discussed in the following sections.

RELATIVE VS. ABSOLUTE COMPARISONS

By comparing a student's overall course performance with that of some relevant group of students, the instructor assigns a grade to show the student's level of achievement or standing within that group. An "A" might not represent excellence in attainment of knowledge and skill if the reference group as a whole is somewhat inept. All students enrolled in a course during a given semester or all students enrolled in a course since its inception are examples of possible comparison groups. The nature of the reference group used is the key to interpreting grades based on comparisons with other students.

Some Advantages of Grading Based on Relative Comparisons

1. Individuals whose academic performance is outstanding in comparison to their peers are recognized.

2. The system is a common one with which most faculty members are familiar. Given additional information about the students, instructor, or college department, grades from the system can be interpreted easily.

Some Disadvantages of Grading Based on Relative Comparisons

1. No matter how outstanding the reference group of students is, some will receive low grades; no matter how low the overall achievement in the reference group, some students will achieve high grades. Grades are difficult to interpret without additional information about the overall quality of the group.
2. Usually the reference group used for comparison should be larger than a single class, especially in courses with enrollments under 40-50 per offering. Instructors who are new to a course have only their first class available for comparison purposes. The performance of that class may not be typical relative to that of future classes.
3. This method encourages competition between students. Though competition can be healthy for many students, it creates unwanted pressures for others. While some are pushed to higher levels of excellence, others are made to feel helpless, hopeless, or even antagonistic toward their peers.

Grades may be determined by comparing a student's performance with specified absolute standards. In this grading method, the instructor is interested in indicating how much of a domain of knowledge or tasks a student knows, rather than how many other students have learned more or less of that domain. A "C" in an introductory statistics class might indicate that the student has minimal knowledge of descriptive and inferential statistics. A much higher achievement level would be required for an "A".

Note that students' grades depend on their level of content mastery; thus the levels of performance of their classmates have no bearing on the final course grade. There are no quotas in each grade category. It is possible that all students in a given class could receive an "A" or a "B" or any other grade.

Some Advantages of Grading Based on Absolute Standards

1. Course goals and standards must necessarily be defined clearly and communicated to the students--instructional planning is enhanced.
2. Most students, if they work hard enough and receive adequate instruction, can obtain high grades. The focus is on achieving course goals, not on competing for a grade.
3. Final course grades reflect achievement of course goals. The grade indicates "what" a student knows rather than how well she or he has performed relative to the reference group.
4. Students do not jeopardize their own grade if they help another student learn course material.

Some Disadvantages of Grading Based on Absolute Standards

1. It is difficult and time consuming to determine what course standards should be for each possible course grade issued.
2. The instructor has to decide on reasonable expectations of students and necessary prerequisite knowledge for subsequent courses. Inexperienced instructors may be at a disadvantage in making these assessments.
3. A complete interpretation of the meaning of a course grade cannot be made unless the major course goals are also known.

STATUS VS. GROWTH

Student's grades may be based on the knowledge and skill they possess at the end of a course compared to their level of achievement at the beginning of the course. Large gains are assigned high grades and small gains are assigned low grades. Students who enter a course with some knowledge of course content are obviously penalized; they have less to gain from a course than does a relatively naive student. The posttest-pretest gain score is more error-laden, from a

measurement perspective, than either of the scores from which it is derived. Though growth is certainly important when assessing the impact of instruction, it is less useful as a basis for determining course grades than end-of-course competence. The value of grades that primarily indicate growth in a college-level course is probably quite low.

ACHIEVEMENT VS. EFFORT

Course grades might represent the amount students learned in a course relative to how much they could be expected to learn as predicted from their measured academic ability. Students with high ability scores (e.g., scores on the Scholastic Aptitude Test (SAT) or American College Test (ACT)) would be expected to achieve higher final examination scores than those with lower ability scores. Then grades are based on comparisons with predicted ability, and an "overachiever" and an "underachiever" may receive the same grade in a particular course, yet their levels of competence with respect to the course content may be vastly different. The first student may not be prepared to take a more advanced course, but the second student may be. When a course grade indicates the amount of effort the instructor believes a student has put into a course, the high ability student who can satisfy course requirements with minimal effort is penalized for an apparent "lack" of effort. Since the letter grade alone cannot communicate such information, the value of ability-based grading does not warrant its use.

FINAL COMMENT

A single course grade should represent only one of the several grading comparisons noted above. To expect a course grade to represent more than one of these meanings is too much of a communication burden. Instructors who wish to communicate more than relative group standing, or subject matter competence or level of effort, must find additional ways to provide such information to each student. Suggestions for doing so are noted near the end of Section III.

III. GRADING AS A PART OF EVALUATION

A distinction should be made between the aspects of performance that an instructor evaluates and the subset of those that are useful for determining course grades. Components or variables that contribute to determining course grades should reflect each student's competence in the course content. The components of a grade should be academically oriented--they should not

be tools of discipline or awards for pleasant personalities or "good" attitudes. A student who gets an "A" in a course should have a firm grasp of the skills and knowledge taught in that course. If the student is merely marginal academically but very industrious and congenial, an "A" grade would be misleading and would render a blow to the motivation of the excellent students in the program. Instructors can give feedback to students on many traits or characteristics, but only academic performance components should be used in determining course grades.

CLASS ATTENDANCE

Students should be encouraged to attend class meetings because it is assumed that the lectures, demonstrations, and discussion will facilitate their learning. If students miss several classes, then their performance on examinations, papers, and projects likely will suffer. If the instructor reduces the course grade because of absence, the instructor is essentially submitting such students to "double jeopardy". For example, an instructor may say that attendance counts ten percent of the course grade, but for students who are absent frequently this may in effect amount to 20 percent. Teachers who experience a good deal of class "cutting" might examine their classroom environment and methods to determine if changes are needed.

CLASS PARTICIPATION

Obviously seminars and small classes depend on student participation to some degree for their success. When participation is important, it may be appropriate for the instructor to use participation grades. In such cases the instructor should keep systematic notes regarding frequency and quality of participation; waiting until the end of the semester and relying strictly on memory makes a relatively subjective task even more subjective. Participation should probably not be graded in most courses, however. Dominating or extroverted students tend to win and introverted or shy students tend to lose. Students should be graded in terms of their achievement level, not in terms of their personality type. Instructors may want to give feedback to students about various aspects of the student's personality but grading should not be the means of doing so.

MECHANICS

Neatness in written work, correctness in spelling and grammar, and organizational ability are all worthy traits. They are assets in most vocational endeavors. To this extent it seems

appropriate that instructors evaluate these factors and give student's feedback about them. However, unless the course objectives include instruction in these skills, students should not be graded on them in the course. A student's grade on an essay exam should not be influenced by his/her general spelling ability, neither should his/her course grade.

PERSONALITY FACTORS

Most instructors are attracted to students who are agreeable, friendly, industrious, and kind; we tend to repel those with opposite characteristics. To the extent that certain personalities may interfere with class work or have limited chances for employment in their field of interest, constructive feedback from the instructor may be necessary. An argumentative student who earns a "C" should have only a moderate amount of knowledge about the course content; the "C" should not reflect the student's disposition directly. The nature of his or her personality should not have direct bearing on the course grade earned.

Instructors can and should evaluate many aspects of student performance in their course. However, only the evaluation information that relates to course goals should be used to assign a course grade. Judgments about writing and speaking skills, personality traits, effort, and motivation should be communicated in some other form. Some faculty use brief conferences for this purpose. Others communicate through written comments on papers or through the use of mock letters of recommendation.

IV. BASIC GRADING GUIDELINES

1. GRADES SHOULD CONFORM TO THE PRACTICE IN THE DEPARTMENT AND THE INSTITUTION IN WHICH THE GRADING OCCURS.

Grading policies of the department, college or campus may limit the grading procedures that can be used and force a basic grading philosophy on each instructor in that administrative unit. Some departments have written statements that specify a method of assigning grades and the meaning of each grade. If such grading policies are not explicitly stated for faculty use, the percentages of A's, B's, C's, D's, and F's given by departments and colleges in their courses may be indicative of implicitly stated or unstated grading policies.

The University regulations encourage a uniform grading policy so that a grade of A, B, C, D, or F will have the same meaning independent of the college or department issuing the grade. In practice grade distributions vary by department, by college and, over time, within each of these units. The grading standards of a department or college are usually known by other campus units. For example, a "B" in a required course given by Department X might indicate that the student probably is not a qualified candidate for graduate school in that field or a related one. Or, a "B" in a required course given by Department Y might indicate that the student's knowledge is probably adequate for the next course. Grades in certain "key" courses may also be interpreted as a sign of a student's ability to continue work in the field. The faculty member who is uninformed about the grading grapevine may unknowingly misjudge a student's potential or misinterpret grading information received. If an instructor's grading pattern differs markedly from others in the department or college and the grading is not being done in special classes (e.g., honors, remedial), the instructor should re-examine his or her grading practices to see that they are rational and defensible (see section VII). Sometimes an individual faculty member's grading policy will differ markedly from that of the department and/or college and still be defensible. For example, the course structure may seem to require a grading plan that differs from departmental guidelines, or the instructor and department may hold different ideas about the function of grading. Usually in such cases, a satisfactory grading plan can be worked out. Faculty new to the University can consult with the department chair for advice about grade assignment procedures in particular courses.

2. GRADING COMPONENTS SHOULD YIELD ACCURATE INFORMATION.

Carefully written tests and/or graded assignments (homework, papers, projects) are important to accurate course grading. Because it is not customary at the university level to accumulate many grades per student in a semester, each grade carries significant weight and should be as accurate as possible. Poorly planned tests and assignments increase the likelihood that grade differences between students will be influenced too much by factors of chance. Some faculty members argue that over the course of a college education, students will receive an equal number of higher-grades-than merited and lower-grades-than merited. Consequently, final GPAs will be relatively correct. However, in view of the many ways course grades are used, each course grade is often significant in itself to the student and others. No evaluation efforts can be expected to be perfectly accurate, but there is merit in striving to assign course grades that most accurately indicate the level of competence of each student.

3. **GRADING PLANS SHOULD BE COMMUNICATED TO THE CLASS AT THE BEGINNING OF EACH SEMESTER.**

By stating the grading procedure at the beginning of a course, the instructor is essentially making a "contract" with the class about how each student is going to be evaluated. The contract should provide the students with a clear understanding of the instructor's expectations so that the students can structure their work efforts. Students should be informed about: which course activities will be included in their final grade; the importance or weight of exams, quizzes, homework sets, papers and projects; and any other factors that might influence their course grade. Students also need to know what method will be used to assign their course grade and what kind of comparison the course grade will represent. By informing students early in the semester about course priorities, the instructor encourages students to study what he or she deems most essential. All of this information can be communicated effectively as a part of the course outline or syllabus.

4. **GRADING PLANS STATED AT THE BEGINNING OF THE COURSE SHOULD NOT BE CHANGED WITHOUT THOUGHTFUL CONSIDERATION AND A COMPLETE EXPLANATION TO THE STUDENTS.**

A common complaint found on students course evaluation forms is that grading procedures stated at the beginning of the course were either inconsistently followed or were changed without explanation. One could look at the situation of altering the grading plan as being analogous to playing a game wherein the rules change unsystematically, sometimes without advance warning to the players. Participation in such games is an extremely difficult and frustrating experience. Students are placed in the unreasonable position of never knowing for sure what the instructor considers important. On those rare occasions when the rules need to be changed, all of the students should be informed (and hopefully be in agreement).

5. **THE NUMBER OF COMPONENTS OR ELEMENTS USED TO ASSIGN COURSE GRADES SHOULD BE LARGE ENOUGH TO ENHANCE HIGH ACCURACY IN GRADING.**

From a decision-making point of view, the more relevant pieces of information available to the decision-maker, the more confidence one can have that the decision will be accurate and appropriate. This same principle applies to the process of assigning grades. If only a final exam score is used to assign a course grade, the adequacy of the grade will depend on how well the test

covered all the relevant aspects of the course content and how typically the student performed on one specific day during a two-hour period. Though the optimum number of tests, quizzes, papers, projects, and/or presentations needed must be course-specific, each instructor must attempt to secure as much relevant data as are reasonably possible to ensure that course grades will represent each student's achievement level accurately.

V. SOME METHODS OF ASSIGNING COURSE GRADES

Various grading practices or methods are used by faculty to assign grades. Some of the more widely used methods and the advantages, disadvantages and fallacies associated with each are described in this section.

Weighting Grading Components and Combining Them to Obtain a Final Grade

Course grades typically are based on a number of graded components (e.g., exams, papers, projects, quizzes) and instructors often weight some components more heavily than others. For example, four combined quiz scores may be valued at the same weight as each of four hourly exam scores (i.e., each quiz is 5% and each hourly exam is 20%). When assigning weights the instructor should consider the extent to which:

each grading component measures important goals;

achievement can be measured accurately with each grading component;

each grading component measures a different area of course content or different set of instructional objectives relative to other components.

The comments in the remainder of this section are based on the assumption that points or scores from several grading components need to be merged to form a composite. Further, course grades are assigned on the basis of the composite score rather than by averaging grades from the separate components. Once it has been decided how much weight each grading component should have, the instructor should ensure that the composite score is actually formed with the desired weights. This task is not as simple as it first appears. An extreme example of weighting will illustrate the potential problem.

Suppose that a 40-item exam and an 80-item exam are to be combined so they have equal weight (50%/50% of the total). We must know something about the spread of scores or variability (e.g., standard deviation) on each exam before adding the scores together. For example, assume that scores on the shorter exam are quite evenly spread throughout the range 10-40, and the scores on the longer exam are in the range 75-80. Because there is so little variability on the 80 item exam, the scores from it will have very little weight in the total score. The net effect is like adding a constant value to each student's score on the 40-item exam; the students maintain essentially the same relative standing that they had on the 40-item exam.

The information appearing in Table 1 demonstrates how scores can be adjusted to achieve the desired weighting before combining them.

	Exam No. 1	Exam No. 2	Total
Number of items	40	80	120
Standard deviation	7.0	3.5	
Desired weight	1	1	
Observed weight	2	1	
Multiplying factor	1	2	
New standard deviation	7.0	7.0	
Actual weight	1	1	

Figure 1. Combining Scores in a Weighted Composite

Exam No. 2 is twice as long as the first, but there is twice as much variability in Exam No. 1 scores (7.0 vs. 3.5) (This is the "observed weight.") The standard deviation tells us, conceptually, the average amount by which scores deviate from the mean of the test scores. The larger the value, the more the scores are spread throughout the possible range of test scores. The variability of scores (standard deviation) is the key to proper weighting. If we merely add these scores together, Exam No. 1 will carry 66 percent of the weight and Exam No. 2 will carry 33 percent of the weight. We must adjust the scores on the second exam so that the standard deviation of the scores will be similar to that for Exam No. 1. This can be accomplished by multiplying each score on the 80-item exam by 2; the adjusted scores will become more varied (new standard deviation = 7.0). The score from Exam No 1 can then be added to the adjusted score from Exam No. 2 to yield a total in which the components are equally weighted. An easier way to accomplish weighting component scores is to transform the raw scores to standard scores, z or T, before applying relative weights.

Converting raw scores to standard scores puts the score distributions of different test components on comparable scales with equivalent means and standard deviations. The item analysis program available from the Exam Service automatically provides T-scores. In addition, the Composite program from the Exam Service can be used by instructors to weight scores, combine scores to form a composite, and assign course grades. Instructors interested in using the Composite program should request a copy of Technical Bulletin #24 from the Exam Service. (Additional readings can be found in Ebel, pp. 252-255; Gronlund, pp. 523-525; Mehrens and Lehmann, pp. 11-13; and Terwilliger, pp. 160-171.

There are several fallacies associated with weighting. First, some instructors believe that if they double the scores from one test and add these scores to a second set of scores, the first test will have twice the weight of the second. Had this been done with the data in Table 2, the first exam would have four times the weight of the second. Another fallacy relates to test length. Some believe that a 100-item test will have twice the weight of a 50 item test if one simply adds scores from the two together. The data in Table 1 illustrates that this notion is inaccurate.

After grading weights have been assigned and combined scores are calculated for each student, the instructor must change the numbered scores into one of five or twelve letter grades. There are several ways of doing this, but some are more appropriate than others. The method chosen must be consistent with the type of comparison the instructor has chosen to give meaning to the grades.

The Distribution Gap Method

This well-known method of assigning test or course grades is based on the relative ranking of students in the form of a frequency distribution, a listing of obtained scores from high to low. The frequency distribution is carefully examined for gaps, several consecutive scores which no student obtained. A horizontal line is drawn at the top of the first gap ("Here are the A's!") and a second gap is sought. The process continues until all possible grade ranges are identified. The major fallacy with this technique is the dependence on "chance" to form the gaps. The gaps are random occurrences because measurement errors (due to guessing, poorly written items, etc.) dictate where gaps will or will not appear. There may be large gaps in a score distribution simply because achievement levels in the group tend to be clustered. But the locations of small gaps here and there throughout the score distribution are due more to random

error than achievement differences. If scores from an equivalent test could be obtained from the same students, the gaps likely would appear in different places. Using the new scores, some students would get higher grades, some would get lower grades, and many grades would remain unchanged. Unless the instructor has additional achievement data to reevaluate borderline cases, many students could see their fate determined as much by chance as by performance.

Grading on the Curve

This method of assigning grades is based on group comparisons and is complicated by the need to establish arbitrary quotas for each grading category. What percent should get As? Bs? Ds? Once these quotas are fixed, grades are assigned without regard to overall quality of performance. The highest ten percent may have achieved at about the same level. But those who "set the curve" or "blow the top off the curve" are merely among the top group; their grade may be the same as that of a student who scored 20 points lower. The bottom five percent may be assigned Fs even though the bottom fifteen percent may be relatively indistinguishable in achievement. Quota-setting strategies vary from instructor to instructor and department to department but seldom carry a defensible rationale. While some instructors defend the use of the normal or bell-shaped curve as an appropriate model for setting quotas, using the normal curve is as arbitrary as using any other curve. It is unlikely that our students' abilities or achievement are normally distributed. Grading on the curve is efficient from an instructor point of view. Therein lies the only merit in the method.

Percent Grading

Though long-standing in use, percent grading in any form is of questionable value. Scores on papers, tests, and projects are typically converted to a percent based on the total possible score. The percent score is then interpreted as the percent of content, skills or knowledge over which the student has command. Thus an exam score of 83 percent means that the student knows 83 percent of the content that is sampled by the test items.

Grades are usually assigned to percent scores using arbitrary standards similar to those set for grading on the curve, i.e., students with scores 93-100 get As and 85-92 is a B, 78-84 is a C, etc. The restriction here is on the score ranges rather than on the number of individuals who can earn each grade. But should the cutoff for A be 92 instead? Why not 90? What sound rationale can be given for any particular fixed cutoff? It seems indefensible (in most cases) to set grade cutoffs that remain constant throughout the course and over consecutive offerings of the course.

It does seem defensible for the instructor to decide on cutoffs for each grading component, independent of the others, so that, for example, the scale for an A might be 93-100 for Exam No. 1, 88-100 for a paper, 87-100 for Exam No. 2, and 90-100 for the final exam.

Some instructors who use percent grading find themselves in a bind when the highest score obtained on an exam is only 68 percent, for example. Was the examination much too difficult? Did students study too little? Was instruction relatively ineffective. Oftentimes, instructors decide to "adjust" scores so that 68 percent is equated to 100 percent. For example, if there were 50 points on the test and 34 was the highest score, each student's percent score would be computed using 34 as the maximum rather than 50.) Though the adjustment might cause all concerned to breathe easier, the new score is essentially the percentage of exam content learned as determined by the highest scoring student. An exam score of 83 no longer means that the student knew 83 percent of the content sampled by the exam.

A Relative Grading Method

Using group comparisons for grading seems appropriate when the class size is sufficiently large (perhaps 40 students or more) to provide a reference group that is representative of students typically enrolled in the course. The following steps describe a widely-used and generally sound procedure:

1. Convert raw scores from each grading component to a standard score (z or T) by using the mean and standard deviation from each respective test, set of papers, or presentation (see Appendix A). Standard scores are recommended because they allow us to view performance on each grading component with a common score scale or standard yardstick. When relative comparisons are to be made, it is not advisable to convert raw scores of components to grades and then average the separate grades. The scores of the separate components should be weighted as described earlier and then combined to form a composite score. Grades are then assigned to composite scores. Because component scores tend to be more reliable (accurate) than component grades, it is preferable to weight the scores rather than the grades. The result should be more reliable composites and course grades.
2. Weight the standard score for each grading variable before combining the standard scores of each student. For example, double both exam standard scores and the standard score for the paper, triple the final exam standard score, and do nothing to the standard score

for the presentation. The respective weights for these variables in the total will then be 20 percent, 20 percent, 20 percent, 30 percent and 10 percent.

3. Add these weighted scores to get a composite (total) score.
4. Build a frequency distribution of the composite scores by listing all obtained scores and the number of students receiving each. Calculate the mean, median, and standard deviation (see Appendix A). Most calculators now available will perform these operations quickly.
5. If the mean and median are similar in value (within a point), use the mean for further computations, otherwise use the median. Let's assume we have chosen the median. Add one half of the standard deviation to the median and subtract the same value from the median. These are the cutoff points for the range of Cs.
6. Add one standard deviation to the upper cutoff of the Cs to find the A-B cutoff. Subtract the same value from the lower cutoff of the Cs to find the D-F cutoff.
7. Use number of assignments completed or quality of assignments or other relevant achievement data available to reevaluate borderline cases. Measurement error exists in composite scores too!

Instructors will need to decide logically on the values to be used for finding grade cutoffs (one-half-, one-third, or three-fourths of a standard deviation, for example). How the current class compares to past classes in ability should be judged in setting standards for each class. When B rather than C is considered the average grade, step five will identify the A-B and C-B cutoffs. Step six would be changed accordingly.

Relative grading methods like the one outlined above are not free from limitations; subjectivity enters into several aspects of the process. But a systematic approach similar to this one, and one that is described in the first class meeting, is not likely to be subject to charges of capricious grading or miscommunication between student and instructor.

An Absolute Standard Grading Method

Absolute grading is a form of assigning grades that is compatible with mastery or near-mastery teaching and learning strategies. The instructor must be able to describe learner behaviors expected at the end of instruction so that grading components can be determined and measures can be built to evaluate performance. Objectives of instruction are provided for students to guide their learning, achievement measures (tests, papers, and projects) are designed from the sets of objectives.

Each time achievement is measured, the score is compared with criteria or standards set by the instructor. Students who do not meet the minimum criterion level study further, rewrite their paper, or make changes in their project to prepare to be evaluated again. This process continues until the student meets the minimum standard established by the instructor. The standards are an important element in the use of this grading method. The following example illustrates how the procedures can be implemented step-by-step:

1. Assume that a test has been built using the objectives from two units of instruction. Read each test item and decide if a student with minimum mastery could answer it correctly. For short answer or essay items, decide how much of the ideal answer the student must supply to demonstrate minimum mastery. Make subjective decisions, in part, on the basis of whether or not the item measures important prerequisites for subsequent units in the course or subsequent courses in the student's program of study.
2. The sum of the points from the above step represents the minimum score for mastery. Next decide what grade the criterion score should be associated with. (Assume for our purposes that the criterion represents the C-B grade cutoff.)
3. Reexamine items which students are not necessarily expected to answer correctly to show minimum mastery. Decide how many of these items "A" students should answer correctly. Such students would exhibit exceptionally good preparation for later instruction.
4. Add the totals from Steps 1 and 3 to find the criterion score for B-A grade cutoff.
5. Each criterion score set in the above fashion should be adjusted downward by 2-4 points, depending on the test length. This adjustment takes measurement error into account. it

compensates for the fact that as test constructors, we may write a few ambiguous or highly difficult items that a well-prepared student might miss due to our own inadequacies. Obviously an adjustment upward could be made if we expect guessing to play a significant part in the test. The adjustment downward simply gives the benefit of doubt to the student, a position to which most faculty likely subscribe.

6. After the exam has been scored, assign "A", "B", and "C or less" grades using the criterion scores. Students who earn "C or less" can be given the opportunity to take a different but equivalent form of the test. A criterion score must be set for this test as described in Step 1. Students who score above the criterion on the second test can earn a "B" at most. Those who fail to meet the criterion on the second test might be examined orally by the instructor for subsequent checks on their mastery.
7. Weight the grades from the separate exams, papers, presentations, and the projects according to the percentages established at the outset of the course. Average the weighted grades (using numerical equivalents, e.g., A=5, B=4, etc.) to determine the course grade. Borderline cases can be reexamined using additional achievement data from the course.

The method described above is not without limitations. The instructor must exercise subjectivity in describing the behaviors that "A" students, for example, must display. Instructors within the same field may not agree on the minimum expectations to require of a "passing" student. Yet this method is unlikely to be labeled as arbitrary when instructors are willing and able to define performance standards in writing and are able to supply a rationale for their judgment. Other methods of determining passing scores or criterion scores are described by Livingston and Zieky (1983).

VI. GRADING IN MULTI-SECTIONED COURSES

Some rather unique grading problems are associated with large multiple-sectioned courses taught by several different instructors, often under the direction and leadership of one head instructor or coordinator. In many of these situations there is a common course outline or syllabus, a common text, and a set of common classroom tests. The head instructor is often concerned about the potential lack of equity in grading standards and practices across the many sections. To promote fairness and equality, the following conditions might be established as part of course planning and monitored throughout the semester by the head instructor:

--The number and type of grading components (e.g., papers, quizzes, exams) should be the same for each section.

--All grading components should be identical or nearly equivalent in terms of content measured and level of difficulty.

--Section instructors should agree on the grading standards to be used (e.g., cutoff scores for grading quizzes, papers, or projects; weights to be used with each component in formulating a semester total score; and the level of difficulty of test questions to be used).

--Evaluation procedures should be consistent across sections (e.g., method of assigning scores to essays, papers, lab write-ups, and presentations).

Though all of the conditions can be addressed in the course planning stage, their implementation may be a more difficult task. Successful implementation requires a spirit of compromise between section instructors and the head instructor as well as among section instructors. Frequent review of instructor practices by the coordinator and constructive feedback to section instructors are needed. The following guidelines contain suggestions for promoting equity in grading across multiple sections:

1. To establish common grading components in each course section, all section instructors should agree at the beginning of the course on the number and kind of components to be used. Agreement should also be reached on the component weighting scheme and final requirements for each course grade (A, B, C, etc.).
2. To encourage instructional adequacy across sections, many coordinators distribute the same course objectives, outlines, lecture notes and handouts to all section instructors. If each instructor is allowed to contribute to the development of common tests, quizzes, or project assignments, the section instructors will become more aware of important course content and the expectations of the coordinator. This awareness will serve to "standardize" section instruction, also.
3. Prior to the administration of an exam, quiz or project, all instructors should agree on established letter grade cutoff scores. The group consensus helps to standardize the administration of grading procedures by reducing the number of "lone wolves" who wish

not to conform to someone else's standards. When relative grading is used, all instructors should agree in advance on the method to be used to assign grades. Ideally, when standard scores are used, they should be computed from the combined scores of all sections rather than on a single section basis.

4. In cases where the grading of particular components is more subjective than objective (i.e., more influenced by personal judgment), organized group practice may help to unify the application of evaluation procedures. For example, coordinators may wish to distribute examples of A, B, or C quality projects to section instructors as models prior to the grading of their own class projects. Or, groups of instructors may wish to practice grading a stack of essay exams by circulating and discussing their individual ratings. Through such group practice the instructors involved can compare their evaluation practices with one another and become more uniform over time.
5. Any grading or evaluation changes made in a particular section should be implemented in all sections. Students should not feel a need to change sections because grading seems easier or more reasonable in another section of the same course.

VII. EVALUATING GRADING POLICIES

Instructors can compare their grade distributions with the grade distributions for similar courses in the same department. Information about grade distributions is available in each department office.

EXAMPLE:

Suppose you taught one section of a 100-level course with 40 students. The course is the first in a three-course sequence that is required in the students curriculum. Your grade distribution turned out to be:

A = 5% B = 20% C = 40% D = 30% F = 5%

When you compare your course grade distribution with that of all of the previous year's sections of the same course, you found the following grade distribution:

A = 22% B = 30% C = 38% D = 9% F = 1%

Because your grade distribution is not consistent with past practice, further investigation is warranted to find out if your particular class was atypical, if your expectations were too high, if the exams upon which the grades were based were too difficult for the course, etc. The fact that your grade distribution does not resemble the grades assigned by your colleagues does not necessarily indicate that your grading methods are incorrect or inappropriate. However, discrepancies that you regard as significant should suggest the need for reexamination of your grading practices in light of department or college policies. The office of the Registrar provides yearly grade distribution reports for departments by semester.

VIII. ASSISTANCE OFFERED BY THE EVALUATION AND EXAMINATION SERVICE (EES)

CONSULTATION

The Exam Service Staff is prepared to discuss grading practices and procedures with faculty who request to do so. The role of the consultants is to explore alternative methods of grading and describe the services available from the office that are designed to assist instructors with evaluating student achievement. Individual consultation and group workshops can be arranged by calling or stopping by the office.

The test scoring and analysis services available from EES provide much of the information needed by faculty to carry out grading procedures. Some of these services can be used even though objective examinations are not used by an instructor.

COMPOSITE

The COMPOSITE system is a service that uses the computer to maintain records of students' test scores throughout a semester. The computer programs used perform much of the mathematics associated with weighting and combining scores. A more complete description of the system is found in Technical Bulletin No. 24, available from the Exam Service

IX. APPENDIX A

TEST STATISTICS

1. Mean Score (\bar{X}) = $\frac{\sum X}{n}$ = $\frac{\text{sum of all scores}}{\text{number of scores}}$

where:

- $\sum X$ = sum of all X
- X = a test score
- n = number of test scores

2. Median Score (Mdn) = The 50th percentile or the score on either side of which half the scores occur after the scores have been ordered.

3. Standard Deviation (SD) = $\frac{\sqrt{-n \sum X^2 - (\sum X)^2}}{n(n - 1)}$

where:

- $\sum X^2$ = sum of all squared test scores
- $(\sum X)^2$ = squared sum of all test scores
- n = number of test scores

NOTE: Many pocket calculators are programmed to compute means and standard deviations. Exam statistics can also be obtained by having objective exams scored and analyzed by EES.

STANDARD SCORES

4. z-score = $\frac{X - \bar{X}}{SD}$

where:

- X = a student's test score
- \bar{X} = mean score
- SD = standard deviation of the scores

5.
$$\underline{\text{T-Score}} = 50 + 10(z)$$

The z- and T-score formulas serve the function of transforming the exam score to a score that has a constant meaning across all different sets of such scores for the same group of individuals. The z-score identifies the number of standard deviation units that a score is above or below the mean. For a z-score of .5, the corresponding exam score is one-half of a standard deviation above the mean. Similarly, a z-score of -0.5 is one-half a standard deviation below the mean. A T-score is simply a converted z-score that eliminates the need for decimal points and negative numbers. A T-score is computed by multiplying a z-score by 10 and adding 50 to the result. Thus, a T-score of 60 represents an exam score that is one standard deviation above the mean, whereas a T-score of 40 is one standard deviation below the mean.

Standard scores (z or T) provide information about a student's performance relative to the performance of the entire class. If we know that Student A received an exam score of 52, we cannot be sure how well Student A performed relative to the rest of the class. However, the information that Student A obtained a z-score of +1.5 (T-score = 65) reveals that the performance was one and a half standard deviations above the class mean, or rather high in comparison to the rest of the class. Standard scores can be obtained with the test analysis and scoring service provided by EES.

X. REFERENCES FOR FURTHER READING

- Dressel, Paul L. & Associates. Evaluation in Higher Education. Boston, Mass: Houghton Mifflin & Co., 1961. Chapter 8, "Testing and Grading Policies"
- Ebel, Robert L. and Frisbie, D.A. Essentials of Educational Measurement (4th ed.). Englewood Cliffs, CA: Prentice Hall, Inc. 1991. Chapter 15, "Grading and Reporting Achievements"
- Frisbie, D.A. Issues in formulating course grading policies. National Association of Colleges and Teachers of Agriculture Journal. 1977, 21, 15-18.
- Frisbie, D.A. Methodological considerations in grading. National Association of Colleges and Teachers of Agriculture Journal. 1978, 22, 30-34.
- Gronlund, Norman E. Measurement and Evaluation in Teaching (4th ed.). New York: Macmillan Publishing, 1981. Chapter 19, "Marking and Reporting."
- Handlin, Oscar and Mary F. The American College and American Culture: Socialization as a Function of Higher Education. Carnegie Commission on Higher Education., New York: McGraw Hill, 1970.
- Livingston, S.A. and Zieky, M.J. Passing Scores. Princeton: ETS, 1982.
- McKeachie, Wilbert J. Teaching Tips (7th ed.) Lexington, Mass: D.C. Heath and Co., 1978. Chapter 17, "The A,B,C's of Assigning Grades," pp. 174-186.
- Mehrens, William A. and Irvin J. Lehmann. Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, 1978.
- Terwilliger, James S. Assigning Grades to Students. Glenview, IL: Scott, Foresman and Co., 1971.