sulting crop loss from sprayer misuse will probably disappear from the operator's mind. He will view himself as a partner in the educational process, one who will train and advise, and make sure that the student is ready for the performance which is to be jointly evaluated by the college supervisor and himself. After this icebreaker, it is easier to add other meaningful tasks to the objectives list, which the employer will take great pride in teaching to his eager-to-learn trainee.

No longer is that $40,000 combine an investment to be protected; it is a tool, a training tool, to be used at the most practical classroom available for the student, his home farm. Adjusting the planter is now a duty that can be taught and assigned, not reserved. Recordkeeping can be taught and shared. The student will begin to develop confidence with understanding, and the owner-operator may begin to think in terms of a partnership.

A sample objective list might appear as follows:

1. **Collect soil samples from fields "A" and "B" and send them to the soil testing laboratory.**
2. **Apply fertilizer to the fields based on soil test recommendations.**
3. **Calibrate and adjust the small grain drill.**
4. **Determine when to plant the oats and plant at least 10 acres without supervision.**
5. **Calibrate the sprayer.**
6. **Add correct amounts of herbicide and water to the sprayer tank so that recommended amounts per acre are followed.**
7. **Spray herbicides on at least 10 acres of oats without supervision.**
8. **Inspect the oats weekly to determine insect and disease problems and any suspected nutritional deficiencies.**
9. **Adjust the combine to minimize harvest losses.**
10. **Combine at least 10 acres of oats without aid.**
11. **Determine yield and profit per acre for the oat (grain and straw) crop, listing the cash flow costs and selling price per bushel.**

Depending on the variety of crops and/or animals encountered in the farming operation, the list could be expanded in many ways. As discussed earlier, it is often useful to wait until the first visit during the employment period to review and add to the original list. Presenting the student and his employer with a list oι more than a dozen initial objectives may appear overwhelming and diminish the excitement of the trainee towards the new venture.

Summary

Since many institutions are now encouraging or requiring their students to engage in work experience prior to graduation, it is essential to assure that the student reap the full benefits from such opportunities. This cannot be ascertained by merely securing a training station for the student and assuming that the employer will take initiative in training the student to be competent in most facets of the business.

Providing meaningful internship requires a supervisor who can successfully outline measurable behavioral objectives which provide the student with the maximum exposure and practical understanding of the job undertaken. Writing objectives which are agreed upon by the student, employer, and supervisor provides not only guidelines by which the employer will direct the student's training, but provides assignments by which trainee performance may be evaluated. The training station becomes more than a place of employment; it develops into a practical classroom, which it should be.

References

1. Anderson, W.A. and A.D. Flowerday. 1973. "Summer Jobs in Ag Careers." **Farm, Ranch and Home Quarterly** 20(2):11.
2. Evans, D.E. 1977. "Penn State's Project on Cooperative Education." **NACTA Journal** XXI(3):13.
3. Gergen, B. 1976. "Diamonds in the Rough." **Successful Farming** 74(11):27.
4. Gunn, T.I. 1974. "Agricultural Business Internship Program." **NACTA Journal** XVIII(4):76-77.
5. LaPrad, R.G. 1977. "Internship Means Obtaining Jobs." **NACTA Journal** XXX(3):14-18.
6. Lewis, D.T. 1973. "A Course in Soil Morphology, Classification and Survey-Objectives, Methods, and Student Response." **Journal of Agronomic Education** 2:80-84.
7. Mager, R.F. 1967. **Preparing Instructional Objectives.** Fearon Publishers, Inc., Palo Alto, Calif.
8. Vogler, D.E., P.E. Curtis and G.W. Wright. 1972. **Here's How: On-Job-Training** - Agricultural Production; Agricultural Products; Agricultural Resources and Forestry; Agricultural Supplies and Services; Agricultural Mechanics; Ornamental Horticulture. The Interstate Printers and Publishers, Inc. Danville, Ill.
9. Zubrick, P. 1976. "An Approach to Curriculum Development." **NACTA Journal** XX(4):22-24.

Methological Considerations In Grading

David A. Frisbie

Abstract

Philosophical, theoretical, and practical issues should be integrated to establish sound evaluation procedures. The five methods of grading most popular among college instructors differ in their philosophical bases and in their appropriateness from both educational and technical standpoints.

It is generally agreed among educators that course planning is essential to the success of instruction. What to teach, how to sequence the content, what materials to use for instructional aids, and what activities to have students accomplish must be considered. An instructor's plan for teaching is shaped by his/her philosophy of education and knowledge of theory and practice regarding instruction. Because evaluation is one component of the instructional process, it is obvious that philosophical, theoretical, and practical issues should contribute to the planning of evaluation procedures.

Once the major philosophical issues discussed in the preceding article have been resolved by the instructor, procedures for grading which are compatible with the philosophy adopted should be established. There is a body of theory and empirical research which should also

Dr. Frisbie is supervisor of Placement and Proficiency Testing, Measurement and Research Division, Office of Instructional Resources, University of Illinois at Urbana-Champaign.

be considered in outlining grading procedures. After grading variables have been identified and evaluated, they must be combined to yield a single symbol to represent course achievement. How should these components be combined? Should they be differentially weighted? How should letter grades be assigned from a distribution of test scores or weighted scores? What are the limitations and advantages of several of the widely-used methods for assigning course grades? These questions will be addressed below in considering the theory, practice, and philosophy which can be integrated to establish sound grading practices.

How Should the Components of Course Grades be Combined?

An illustration from a course in beef production will be used to demonstrate the process of combining several grading variables to form a composite score. Assume that the following course requirements are to be used as a basis for determining the course grade: a) an hourly essay exam covering the principles of feeding and management, b) an hourly essay and objective exam covering financial aspects of production, c) a term paper relating to diseases and parasites, d) a demonstration or presentation on the topic of breeding, and e) a comprehensive objective final exam. Let us further assume that the instructor uses a norm-referenced approach to grading, i.e., a student's performance is compared with that of his classmates and his relative class standing determines his grade.

How should the exams, papers, and presentation be weighted in arriving at a grade in Beef Production? The first consideration should be the validity of each component, the extent to which each is a measure of important course objectives. The demonstration might be weighted less than the paper, for example, because its brevity restricts it to only a portion of the concepts, principles, and problems of breeding which the course is intended to address. Reliability, the extent to which achievement is measured accurately or with minimal error, is another consideration. The subjectivity generally inherent in grading term papers is a source of measurement error. Scores or grades on papers might carry less weight than objective exam scores for this very reason. The same rationale might be used in deciding on the weight to assign scores on an essay examination. Because instructors vary in their ability to score essays or papers accurately, we should expect variations between instructors in how much weight is assigned to such components in their courses.

The uniqueness of each grading variable should be considered in assigning the component weight. Here uniqueness refers to the objectives which are being measured by each component. Two exams covering distinctly different content objectives are unique; an exam and a paper which both cover diseases and parasites are less unique. To the extent that components are measures of the same objectives they are redundant measures and their combined weight should be assessed. The first four measures in our illustration seem unique; but the final exam is somewhat redundant with each because it is comprehensive, covering the entire course content.

On the basis of validity, reliability, and uniqueness, the components in our illustration are weighted as follows: Hourly 1, 20 percent; Hourly 2, 20 percent; Paper, 20 percent; Presentation, 10 percent; and Final Exam, 30 percent. We obviously cannot agree or disagree with these weights without knowing more about the demands and expectations associated with each component.

The next problem facing the instructor is to insure that the weights assigned above are actually used when the components are combined. Though this appears simple to achieve, it is a knotty problem which is avoided by many instructors.

An extreme example of weighting will exaggerate the impact while illustrating the procedure. Suppose that a 40-item exam and an 80-item exam are to be combined so they have equal weight (50 percent-50 percent) in 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. Assume that scores on the shorter exam are quite evenly spread throughout the range 10-40, and the scores on the other are in the range 75-80. Because there is so little variability on the 80-item exam, if we merely add each student's scores together, the spread of scores in the total will be very much like the spread of scores observed on the first exam. The second exam 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.

The information appearing in Exhibit 1 will be used to demonstrate how scores can be adjusted to achieve the desired weighting before combining them. Exam No. 2 is twice as long as the first, but there is twice as much variability in Exam No. 1 scores. (The standard deviation tells us, conceptually, the average amount by which scores deviate from the test mean. The larger the value, the more scores are spread out through 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 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 two; the adjusted scores will become more varied (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. (A practical solution to combining several weighted components is to first transform raw scores to standard scores, z or T, before applying relative weights and adding.) (Additional reading can be found in Ebel, pp. 348-51; Gronlund, pp. 523-5; Mehrens and Lehmann, pp. 600-1; and Terwillinger, pp. 160-71.)

Exhibit 1. Combining Scores in a Weighted Composite

	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	

Some Methods of Assigning Course Grades

Several popular grading methods or practices will be examined below to identify some of the advantages, disadvantages, and fallacies associated with each.

The Distribution Gap Method. This widely-used method of assigning test or course grades is based on the relative ranking of students in the form of a frequency distribution. The frequency distribution is carefully scrutinized for gaps, several consecutive scores which have zero frequency. 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 (A-F) are identified. The major fallacy with this technique is the dependence on "chance" to form the gaps. The gaps are random because measurement errors (due to guessing, poorly written items, etc.) dictate where gaps will or will not appear. If scores from an equivalent test could be obtained from the same group, the gaps would likely appear in different places. The implication obviously is that 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 more by chance than performance.

Grading on the Curve. The norm-referenced basis for this type of grading is complicated by the need to establish arbitrary quotas for each grade category. What percent should get A's? B's? D's? Once these quotas are fixed, grades are assigned without regard to level of performance. The highest ten percent may get A's though the next ten percent may have achieved at about the same level. 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 F's though the bottom fifteen percent may be relatively indistinguishable in achievement. Quota-setting strategies vary from instructor to instructor and department to department and seldom carry a defensible rationale. Grading on the curve is efficient from an instructor point of view; therein lies the only merit in the method. (See Terwillinger, pp. 75-8, for further reading.)

Percent Grading. Though many names and alternate techniques are associated with it, the long-standing use of percent grading in any form is questionable. 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 which is represented by the test items. (The test items themselves represent only a sample of the universe of content.) The validity of such a conclusion weighs heavily on both the skill of the test constructor and an adequate definition of the universe of content. The above interpretation of percent scores depends on what we have called absolute standards of performance.

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 A's 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. Should the cutoffs for an A be 94 instead? Why not 90? What sound rationale can be given for any particular cutoff? It seems indefensible in most cases to set grade cutoffs that remain constant throughout the course and several consecutive offerings of the course. It does seem defensible for the instructor to decide on cutoffs for each grading variable, independent of the others, so that 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. Further comments on this type of procedure will be made in a later section.

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. (Assume there were 100 points on the test and 68 was the highest score. Divide all scores by 68 rather than 100 to arrive at "adjusted" percent scores.) Though the adjustment might cause all concerned to breath easier, the new score is essentially uninterpretable in terms of the universe of content the 100-item exam represents.

A Relative Grading Method. Norm-referenced grading seems appropriate for many situations in which the procedures for absolute grading seem impractical or economically infeasible. If the class is large (perhaps 35 students or more) a broader reference group may not be needed. The following steps describe a widely-used and generally sound procedure:

1. Convert raw scores on each exam to a standard score (z or T) using the mean and standard deviation from each respective test, set of papers, or groups or presentations. Do not convert raw scores to grades and average the separate grades. The blood, sweat, and tears shed in trying to distinguish between achievement levels will be lost; differences will melt together as students are forced into a few broad categories.
2. Weigh each grading variable before combining the standard scores. 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 will then be 20 percent, 20 percent, 20 percent, 30 percent and 10 percent.

3. Add these weighted scores to get a composite or total score.
4. Build a frequency distribution and calculate the mean, median, and standard deviation. (Most calculators now available will perform these operations quickly.)
5. If the mean and median are "similar" in value use the mean from this point. 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 C's.
6. Add one standard deviation to the upper cutoff of the C's to find the A-B cutoff. Subtract the same value from the lower cutoff of the C's 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). The entering characteristics of typical classes should be assessed and the typicalness of each class should be judged in setting standards. 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 several aspects of the process. But a systematic approach similar to this one which is thoroughly described in the first class meeting is less subject to charges of capricious grading and miscommunication between student and instructor. (See Terwillinger, pp. 78-97, for a variety of applications of this method.)

An Absolute Grading Method. Absolute grading is the only form of assigning grades which 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 variables can be determined and measures can be built to evaluate performance. Objectives of instruction are provided for students to guide their learning, and achievement measures (tests, papers, and projects) are designed from the sets of objectives.

Each time achievement is measured, the score is compared with some criterion or standard set by the instructor. Students who do not meet the criterion level are recycled, i.e., they 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 standards established by the instructor. The standards are obviously the key to the success 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 have correct to demonstrate minimum mastery. The instructor's subjective decisions should be made, 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 students' programs of study.

2. The sum of the points from the above step represents the minimum criterion score for mastery. The instructor must decide what grade the criterion score should be associated with. (Assume for our purposes that the criterion represents the C-B 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. (This step could be done concurrently with Step 1.)

4. Add the totals from Steps 1 and 3 to find the criterion score for the B-A grade cutoff.

5. After the exam has been scored, assign "A", "B", and "C or less" grades using the criterion scores. Students who earn "C or less" should be given a different but equivalent form of the test within two weeks. A criterion score must be set for this test as described in Step 1. Students who score above the criterion can earn a "B" at most. Those who fail to meet the criterion on the second testing might be examined orally by the instructor for subsequent checks on their mastery.

The grades from the separate exams, papers, presentations, and projects should be weighted according to the percentages established by the instructor at the outset of the course. The weighted grades are then averaged (using numerical equivalents, e.g., A = 4, B = 3, etc.) to determine the course grade. Borderline cases can be reexamined using additional achievement data from the course. (Additional detail can be found in Terwillinger, pp. 26-37.)

An absolute grading method like that outlined above requires a flexible approach to instruction, careful planning, and time for developing equivalent measures. Additional instructor time is needed to readminister exams, reread papers, and reevaluate projects. The "loss" in time in usually offset by the gain in the quality of student performance as a group, the increased sense of accomplishment experienced by students and instructor, and an increase in student-instructor interaction. These assets are admittedly more a product of the instructional strategy than the grading method. But the two go hand-in-hand.

Conclusions

A weighting of grading components is generally achieved by assessing the validity, reliability, and uniqueness of each variable. The variability of the scores (standard deviation) must be examined before several weighted variables are combined. The mere adding of scores will not always yield the desired weights.

The five methods of grading described differed in their philosophical bases and in ther appropriateness from both educational and technical standpoints. There is no "right" method of assigning grades as long as educators differ philosophically about what a grade should mean. The relative grading method described is sound and attractive for those who use a norm-referenced approach. Those who prefer absolute standards should find merit in the absolute grading method described. Some variation in methods of grading can be tolerated as long as those methods are logically and educationally defensible.

References

Ebel, Robert L. 1972. **Essentials of educational measurement.** Prentice-Hall, Englewood Cliffs, NJ.

Gronlund, Norman E. 1976. **Measurement and evaluation in teaching.** Macmillan Publishing, New York, NY.

Mehrens, William A. and Irvin J. Lehmann. 1973. **Measurement and evaluation in education and psychology.** Holt, Rinehart and Winston, New York, NY.

Terwillinger, James S. 1971. **Assigning grades to students.** Scott Foresman, Glenview, IL.

The Relationship of Carrel Use With Subject Matter Taught, Student Background, and Grades in an Autotutorial Crops Lab.

A.W. Burger and R.D. Seif

Abstract

No significant correlation between carrel times and/or total time spent in the laboratory with final course grades was found during four years (1972-1975) of autotutorial crop science laboratory teaching. Urban students spent either more carrel time or total laboratory time than did farm students in completing studies on the identification of common legumes and grasses as well as insects. No significant differences in either carrel or total laboratory times were observed between urban and farm students on studies of: (a) crop or weed seed identification, (b) corn and soybean seedling emergence, (c) basic genetics, (d) classification of flowering plant families, and (e) common crop diseases. Females spent more total laboratory time than males, but less carrel time in completing some objectives in crop science. Freshmen, students with no part-time employment, and students enrolled in non-engineering-mechanization curricula spent either more carrel or total laboratory time in completing some objectives than did sophomores and upper classmen, students with part-time employment, and engineering mechanization students, respectively. In 8 of 10 study units, student carrel times exceeded unit tape times.

Introduction

Autotutorial study by college students is not new (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14). An examination and/or evaluation of this system of teaching is important in order that such a program is conducted as efficiently as possible to promote learning.

The makeup of our University of Illinois college crop science classes is not stable from one year to the next. Five years ago, most of the students in our crop science course came from the farm or they had a farm background. In the spring semester of 1976, 40 percent of our crop science enrollees were from urban areas. Many more female students are majoring in Agronomy. Because of this ever changing makeup of our classes, it becomes necessary to adapt autotutorial materials so that urban as well as rural, female as well as male students can achieve equally well the learning objectives programmed for credit in crop science courses.

The total amount of student time spent in independent study carrels is important in the planning of future learning units. The alert instructor must be concerned constantly with whether there is enough time for all students with varied backgrounds to complete the lesson in a given subject matter within the laboratory period allotted for a given laboratory experience. Units may be too long, too challenging, or too tiring.

This paper examines the possible relationship of carrel use times with: (a) the many subject matter disciplines quite typically taught in a beginning crop science laboratory course at many institutions, (b) student backgrounds, and (c) final grades achieved in the course. It is hoped that the method used in measuring this possible relationship of carrel use times with the many factors of the student's background and environment will be useful to others in future planning and programming of autotutorial instruction.

Methods and Materials

A simple correlation analysis was made to test the possible relationship of carrel times and/or total laboratory times with final course grades of enrollees in the introductory crop science course during seven semesters, spring 1972 - spring 1975, at the University of Illinois. In this study, carrel times are defined as the actual amount of time in minutes consumed by a student in the study

Burger is professor of Agronomy and Self is professor of Biometry-Agronomy, University of Illinois at Urbana-Champaign.