Wanted: Assessment-based grading methods that support student engagement

David M. Keil, Framingham State University (dkeil@framingham.edu)

Assessment Institute, Indianapolis, Indiana, October 20, 2014

Abstract. Grades are usually a point of friction, not connection, between us and our students. How can we turn this around? Can we use grades to help engage the students with what really matters to us? Grading mostly on course learning objectives may help call students' attention to these desired *outcomes*. In contrast, the normal focus is on grades and on tools of evaluation. As we evaluate student work, we can name what matters to us: attainment of course objectives, knowledge of course concepts, contribution, collaboration, problem solving, and critical thinking. Focusing on student engagement helps us to frame our communications. Giving students multiple opportunities to show attainment of objectives enables us to raise our standards.

1. Introduction

A couple of years ago, as I went over my plans for grading my courses, I thought about how I was presenting my grading policies. My syllabi had a section, "Grading," which listed what counted, such as quizzes, exams, assignments, and attendance, and the percentages that each represented in the final grade.

It occurred to me that grades always seem to be the main concern of most students. So what was I saying about my values? I was telling them, in effect, that I valued quizzes, exams, assignments, and attendance, and that those were the price for obtaining the grades that they valued. My list was like the price list on a menu or in a newspaper ad supplement.

But the advertisement wasn't true! I don't teach because I value quizzes, exams, assignments, and attendance. I teach because I love my field and want to share it. What I value is that the students learn about my field and its practices. Quizzes, exams, assignments, and attendance are just tools for helping and tracking this learning.

The unstated falsehood in education is that grades are the payment that students get for writing good papers and taking quizzes. There's even bargaining over grades, as if the school were a flea market.

Grades have most students' highest attention from the first to the last day of the semester. I decided to use the obsession with grades to point to what I value, rather than

to the tools I use for evaluation. I decided to change how I present grading in my syllabi so as to reflect my values and to help make them known to students.

How can we use our grading role, and the way we talk about grades, to turn students somewhat away from their concern about grades and rankings, and somewhat toward the investigation that we hope they will put their efforts to?

2. A system of grading

Below is a new way to present grades. Assessment of student learning, and grading, may be broken into two components: what the student is shown to have learned, and what the student has contributed to the learning of others. I break it down 60-40. The system is based on two assumptions: that we can measure learning, and that students' contribution to the learning of others is measurable, worth counting, and related to their own learning.

I assess learning mostly according to objectives, which I subdivide into "core" and "non-core", but partly according to performance on multiple-choice quizzes, and end-of-semester summary quizzes. I label these for my students according to what I value (learning), not according to the instrument by which I assess learning.
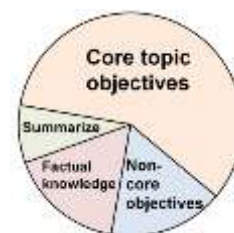
Most core and non-core objectives I assess using in-class quiz answers to questions that support an objective; in some cases, only out-of-class work can validly assess an objective. Since what matters is what a student has learned by the end of the semester, I give multiple opportunities to show attainment of objectives.

A student's contribution to the classroom community can be measured in various ways: physical presence; leading discussions of problems or solutions; turning in exercises in a timely way; documented participation in groups; or other written work, including taking quizzes on time.

I am asking students to turn in their exercises on a learning outcome *before* they answer a quiz question that assesses the outcome. The rationale is that you play in a game *after* you've put in time at practice.

Administratively, it's possible to allow second and third chances only if students make reasonable efforts on the first opportunity; students can't expect to turn in all their work and take all their quizzes only at the end of the semester, for full credit. I have seen some success with second and third chances and some failure.

I don't charge a "penalty" for failure to turn in work on time or failure to take quizzes the first time. I want students to answer quiz questions only after they have studied and are ready. We can acknowledge on-time work, however, as part of the evaluation of the student's contribution to the learning community.

3. Grading according to outcomes

Each question is written so as to assess a course objective, such as "2b. Prove correctness of an algorithm, using invariants." The course objectives come from a list of desired outcomes for courses, adopted by my department. Each objective describes a capacity or skill that a student is to be able to demonstrate by the end of the semester. Some objectives are course-wide; others are topic specific.

I use a rubric (below) that evaluates solutions from 0 to 4, rather than 0 to 100. What seems important is to have a uniform scale (Figure 1) for evaluation of the quality of student answers.

Code	Meaning
6	Solution of rare quality surpassing requirements.
5	Solves problem thoroughly and accurately; applies relevant concepts adeptly, showing mastery of objective.
4	Mostly successful solution with good application of concepts. Strong support for claim of success with objective.
3	A fair-quality solution with omissions or errors. Generally successful application of concepts.
2	A solution that shows some grasp of relevant concepts, meeting minimum standards for the objective.
1	Unsuccessful answer showing some understanding of problem.
0	No answer or irrelevant answer

Figure 1: Rubric for scoring student answers

In the past, I allocated "points" to quiz and assigned problems, so that the grading decision was how many points out of, say fifteen, to give to a student's answer. A good answer might get twelve points. A good answer to a five-point question might get four points. What is the precise meaning of four or twelve to the student? It is likely to be unclear. But students can remember the essence of the rubric: a 6 means "extraordinary," 5 means "excellent," 4 means "fine", 3 means "OK", 2 means "barely successful," and 1 or no-score are requests to try again for success.

For assessment, scores on quiz questions may be collected in a series of spreadsheet columns, one per course objective. An average for a given objective over all students yields assessment information on how well the students attained the objective. An average for a given student over all objectives yields data on how well the student attained all objectives. Objectives may be categorized for importance; e.g., "essential," "priority," or "challenge."

Students may track scores for learning objectives posted online. Thus student attention is drawn not to quiz and assignment artifacts, but to the actual objectives of the course.

If average scores on an objective are lower than those on another objective, I may give it more attention during the semester to present the material more thoroughly. Alternatively, the objective may be deemed hard to attain in the time available, and the scores may be scaled up.

Course-wide objectives may include the following:
- Solve a problem as a member of a team
- Present a short talk in the classroom
- Write a short documented research paper
- Participate in the classroom process

This system of grading may be extended to enable students to show attainment of objectives at several times in the semester, with at least three opportunities offered per objective. The *max* function in the Excel spreadsheet application supports this best-performance notion.

Translating scores from a 0..6 range to a 0..100 range, for grading on an A..F scale, requires a spreadsheet formula.

Allowing multiple opportunities to succeed supports use of the entire semester as an assessment period. I encourage students to correct their assignments and research work for credit after submission. This departs from what is described as "most students' expectations of 'produce a product one time only and receive a grade' " (Taylor). In this model, which students learn in school, the student is made to appear as a vendor and the teacher as a customer.

4. Criteria-based vs. norm-based grading

Criteria-based grading may be distinguished from *norm-based grading*. In norm-based grading, also called grading on the curve, students are graded by relative

rankings rather than by criteria applied to them as individuals.

Four models of criteria-based grading include those based on achievement of course objectives (my grading is in this category); overall achievement according to score totals (the most common model); patterns of achievement; and specified qualitative attributes (Sadler, 2005).

A page at the assessment site at Carnegie Mellon University (*www.cmu.edu*) describes how to use grading data in assessment of student learning. It requires decomposing grades "into the components that are indicators of learning outcomes and those that are indicators of other behaviors." (My grading scheme, in Section 2 above, decomposes grades into assessment of learning outcomes and student contributions to the classroom community.) Second, "separate grades or sub-scores would have to be computed for the major components of knowledge and skills."

This kind of assessment, the CMU site points out, can only be done by the course instructors themselves. Thus, for professors to take the initiative to tie their grading to learning outcomes is in effect a way to empower themselves in the necessary assessment of student learning.

A paper that describes European efforts to standardize systems of higher education, including grading systems, points out that standardized criteria, and even multi-level grading schemes, are of limited usefulness, as opposed to pass/fail systems (Dahlgren et al, 2009). "A multi-step grading system," it asserts, "steers the students towards focusing more on the assessment task itself when planning their studies than on learning." Moreover, a multi-step system is less likely to encourage cooperation among students.

Significantly, the paper notes that "grading systems are among the most influential factors determining the quality of higher education."

5. Raising standards by offering multiple opportunities to succeed

It is common to assume that the normal curve, below, fairly describes success and failure of learning. It is assumed that the majority of students in any environment achieve at an acceptable, but not high-quality, level, and that their work should be evaluated as "C". In fact, "C" is considered to be a fair label for average work, by definition.

Factually, the normal or Gaussian distribution depicts only relative statistical constructs. It is quite possible for any proportion of students in a class to learn at a high-quality level or at a failing level. If no one in a computer-programming class learns by a certain time to write a program, then even those who have attained the 99th percentile by some measure have failed to learn the skill by that time.

If we let students know that success or failure in an endeavor is measured by multiple trials, if necessary, then we are in a position both to encourage those who learn at a slower speed than others and to insist that all learn certain minimal skills as a condition for certified success. *We are in a position to raise standards above the possibly low ones that we may be enabling with our norm-based grading.*

6. Power relationships and student engagement

The mention of collaborative learning in relation to grading raises the question of the relative power of teachers and students, which is most strikingly expressed in the grading relationship. As an alternative to a teacher-centered or student-centered environment, Parker Palmer advocates for a "community of truth" that places the subject of study at the center of the learning environment (Palmer, 2007). His work is part of an influential body of study that challenges the traditional hierarchical understanding of education.

A significant early contribution to this literature is that of Paulo Freire, who suggested the "problem posing" notion of education in place of the "banking" concept. In the banking model, the teacher is the active knower and provider; the students are passive consumers of knowledge. Instead, according to Freire, teachers "must be partners of the students in their relations with them" (Freire, 2009).

The hierarchical model of education sees the teacher's grading role as selection; "to pick the best and the brightest," as Ken Bain summarizes the views of the "traditional" college teachers interviewed in his work.

In this model, which William Glasser characterizes as "boss management," norm-based grading methods are used, "grading on the curve" (Glasser, 2009). Under this system, a predetermined number of students in a group receive *A* grades, a larger number *B*s and *C*s, and a small number receive bottom grades.

Students are in effect pitted against each other, reproducing the hierarchy of teacher-student with an enforced hierarchy among students. There is no community of truth but rather a battle to survive, in which some students must necessarily fail regardless of their efforts or achievements.

If this has an aim of promoting learning effort, the underlying strategy is to rely on the grading power of the teacher to stimulate frightened students to strive not to come out at the bottom.

A recent discussion at Harvard University indicates that some academics consider the curve to be essential. When it was learned that the most common grade at

Harvard was *A*, a longtime professor lashed out publicly with a claim of "grade inflation" (Anderson, 2013). Notably, the news report makes no reference to any claimed lowering of standards for achieving an *A*; only claims that *A* must not be the most frequent grade.

The equivalent method of management, used in industry, has gone under the names "stack ranking," "vitality curve," and "rank and yank." Promoted by Jack Welch of General Electric, this method required selection of a small fixed proportion of employees from each unit as high achievers, meriting bonuses, and another fixed proportion at the bottom worthy only of dismissal.

Microsoft Corporation adopted this system in 2006 and abandoned it in 2013, announcing then a restructuring aimed at improving cooperation among employees. The competitive system had "created an incentive for star Microsoft workers to avoid working with other stars since they knew it could hurt their chances of getting a top rating when it came time for employee reviews" (Wingfield, 2013).

When students are concerned above all about competing with other students for grades, even about failing due to the success of other students, then they may avoid collaboration and may be distracted from learning.

7. Grading group work

If we want students to help each other learn, then we need to provide forms for group work. Simply assigning group projects, for which all students will receive the same grade, appears to invite conflict over division of labor, and dissatisfaction over unfair division of labor, rather than generating collaboration.

Other strategies suggested include requesting:
- minutes of group meetings;
- individual peer evaluations of contributions;
- students' assessments of their own contributions to the group (Erickson et al, 2006)

The request that students assess their own contributions entails reflection on their own learning – a value in itself.

This source included a form suitable for peer evaluations, listing ways that a group member could contribute to the common effort.

8. Evidence that attention to grades distracts from learning

A central conclusion of Bain is that students' *intrinsic* motivation, the desire to learn for its own sake, is the key to their engagement and their learning. This is supported by research reported by Alfie Kohn on the distracting and demotivating effects of rewards and punishments in the workplace, in child raising, and in schools (Kohn, 1999).

Yes, when we say that we will submit grades at the end of the semester, and that grades will be based on work done, some students who might not otherwise have done so will come to class and even pick up a pencil. But will they engage with the subject? Or is it only *curiosity* that can motivate students to deep learning, to problem solving and critical thinking?

9. Conclusion

One way to begin may be to change how we name what we value, when we write in our syllabi, under the heading, "Grading," lists of what counts in our courses.

With our grades, and our presentation of grading, we communicate our values. If we value learning and contribution, we have the option to say so.

Bain and Kohn explain how rewards and punishments don't motivate deep learning. But we do have the option to use our students' focus on grades to turn their attention to learning as what we value.

Learning requires failure and recovery. Therefore our grading and our learning assessment can take account of student learning later in the semester, after failure earlier in the semester.

In a learning environment that is collaborative, we need to assess group work. Methods for this are likely to be an area of conversation in support of student engagement.

References

Travis Anderson, Nicholas Jacques, and Todd Feathers. Harvard Professor says grade inflation rampant. *Boston Globe*, December 14, 2013.

Ken Bain. *What the Best College Teachers Do.* Harvard University Press, 2004.

Eberly Center for Teaching Excellence. Grading Versus Assessment of learning outcomes: What's the difference? Carnegie Mellon University. *http://www.cmu.edu/teaching/ assessment*.

Lars Owe Dahlgren, Andreas Fejes, Madeleine Abrandt-Dahlgren and Nils Trowald. Grading systems, features of assessment and students' approaches to learning. *Teaching in Higher Education 14*:2 (2009), pp. 185-194.

B. Erickson, C. Peters, D. Weltner. *Teaching First-Year Students*. Jossey-Bass, 2006.

Paulo Freire. *Pedagogy of the Oppressed.* Continuum, 2009.

William Glasser. *The Quality School: Managing Students without Coercion*. Harper, 1998.

Alfie Kohn. *Punished by Rewards*. Houghton Mifflin, 1999.

Parker Palmer. *The Courage to Teach*. Jossey-Bass, 2007.

Peter Taylor. More assessment, less grading. *http://www.cct.umb.edu/ moreassessment.html*..

Nick Wingfield. Microsoft abolishes employee evaluation system. *New York Times*, November 13, 2013.