

## Feature

### *Approaches to Biology Teaching and Learning*

## Teaching More by Grading Less (or Differently)

Jeffrey Schinske\* and Kimberly Tanner†

\*Department of Biology, De Anza College, Cupertino, CA 95014; †Department of Biology, San Francisco State University, San Francisco, CA 94132

### INTRODUCTION

When we consider the practically universal use in all educational institutions of a system of marks, whether numbers or letters, to indicate scholastic attainment of the pupils or students in these institutions, and when we remember how very great stress is laid by teachers and pupils alike upon these marks as real measures or indicators of attainment, we can but be astonished at the blind faith that has been felt in the reliability of the marking systems.

—I. E. Finkelstein (1913)

If your current professional position involves teaching in a formal classroom setting, you are likely familiar with the process of assigning final course grades. Last time you assigned grades, did you assign an “E,” “E+,” or “E–” to any of your students? Likely you assigned variations on “A’s,” “B’s,” “C’s,” “D’s,” and “F’s.” Have you wondered what happened to the “E’s” or talked with colleagues about their mysterious absence from the grading lexicon? While we often commiserate about the process of assigning grades, which may be as stressful for instructors as for students, the lack of conversation among instructors about the mysterious omission of the “E” is but one indicator of the many tacit assumptions we all make about the processes of grading in higher education. Given that the time and stress associated with grading has the potential to distract instructors from other, more meaningful aspects of teaching and learning, it is perhaps time to begin scrutinizing our tacit assumptions surrounding grading. Below, we explore a brief history of grading in higher education in the United States. This is followed by considerations of the potential purposes of grading and insights from

research literature that has explored the influence of grading on teaching and learning. In particular, does grading provide feedback for students that can promote learning? How might grades motivate struggling students? What are the origins of norm-referenced grading—also known as curving? And, finally, to what extent does grading provide reliable information about student learning and mastery of concepts? We end by offering four potential adjustments to our general approach to grading in undergraduate science courses for instructors to consider.

### A BRIEF HISTORY OF GRADING IN HIGHER EDUCATION

It can be easy to perceive grades as both fixed and inevitable—without origin or evolution . . . Yet grades have not always been a part of education in the United States.

—Schneider and Hutt (2013)

Surprisingly, the letter grades most of us take for granted did not gain widespread popularity until the 1940s. Even as late as 1971, only 67% of primary and secondary schools in the United States used letter grades (National Education Association, 1971). It is therefore helpful to contextualize the subject to appreciate the relatively young and constantly changing nature of current systems of grading. While not an exhaustive history, the sections below describe some of the main developments leading to the current dominant grading system.

#### *Early 19th Century and Before*

The earliest forms of grading consisted of exit exams before awarding of a degree, as seen at Harvard as early as 1646 (Smallwood, 1935). Some schools also awarded medals based on competitions among students or held regular competitions to assign seats in class (Cureton, 1971). Given that universities like Yale and Harvard conducted examinations and elected valedictorians and salutatorians early in the 18th century, some scale of grading must have existed. However, the first official record of a grading system surfaces in 1785 at Yale, where seniors were graded into four categories: *Optimi*, second *Optimi*, *Inferiores*, and *Periores* (Stiles, 1901, cited by Smallwood, 1935). By 1837, Yale was also recording

DOI: 10.1187/cbe.CBE-14-03-0054

Address correspondence to: Jeffrey Schinske (schinskejeff@deanza.edu).

© 2014 J. Schinske and K. Tanner. CBE—Life Sciences Education © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

student credit for individual classes, not just at the completion of college studies, using a four-point scale. However, these “merit marks” were written in code and hidden from students (Bagg, 1871).

Harvard and other schools soon experimented with public rankings and evaluations, noting that this resulted in “increasing [student] attention to the course of studies” and encouraged “good moral conduct” (Harvard University, 1832). Concerned that such public notices would inspire competition among students, which would distract from learning, other schools used more frequent, lower-stakes “report cards” to provide feedback on achievement (Schneider and Hutt, 2013). In 1837, at least some professors at Harvard were grading using a 100-point system (Smallwood, 1935). During this same period, William and Mary placed students in categories based on attendance and conduct. The University of Michigan experimented with a variety of grading systems in the 1850s and 1860s, including various numeric and pass/fail systems (Smallwood, 1935). Still, many schools at this time kept no formal records of grades (Schneider and Hutt, 2013).

### *Late 19th Century and 20th Century*

With schools growing rapidly in size and number and coordination between schools becoming more important, grades became one of the primary means of communication between institutions (Schneider and Hutt, 2013). This meant grades needed to have meaning not just within an institution but also to distant third parties. A record from 1883 indicates a student at Harvard received a “B,” and in 1884, Mount Holyoke was grading on a system including “A,” “B,” “C,” “D,” and “E.” Each letter corresponded to a range of percentage scores, with lower than 75% equating to an “E” and indicating failure. Mount Holyoke added an “F” grade (for failing) to the scale in 1898 and adjusted the percentages relating to the other letters (Smallwood, 1935). This appears to be the initial origin of the “A”–“F” system familiar to most faculty members today, albeit including an “E” grade. By 1890, the “A”–“E” system had spread to Harvard after faculty members expressed concerns regarding reliably grading students on a 100-point scale. Still, grading was not always done at schools and grading systems varied widely (Schneider and Hutt, 2013).

By the early 1900s, 100-point or percentage-based grading systems were very common (Cureton, 1971). This period also saw an increased desire for uniformity in grading, and many expressed concerns about what grades meant from one teacher or institution to the next (Weld, 1917). Numerous studies of the period sought to understand and perfect grading systems (Cureton, 1971). Grading on a 100-point scale was found to be highly unreliable, with different teachers unable to assign consistent grades on papers in English, math, and history (Starch, 1913). Researchers felt that getting away from a 100-point scale and grading into only five categories (e.g., letter grades) could increase reliability (Finkelstein, 1913, p. 18). While it is unclear exactly when and why “E” grades disappeared from the letter grade scale, it seems possible that this push to use fewer categories resulted in an “A”–“F” scale with no “E” (“F” being retained, since it so clearly stood for “fail”). Others have conjectured that “E” was removed so students would not assume “E” stood for “excellent,” but whatever the reason, “E’s” apparently disappeared by the 1930s (Palmer, 2010).

As research on intellectual ability appeared to show that, like other continuous biological traits, levels of aptitude in a population conformed to a normal curve, some experts felt grades should similarly be distributed according to a curve in a classroom (Finkelstein, 1913). Distributing grades according to a normal curve was therefore considered as a solution to the subjective nature of grading and a way to minimize interrater differences in grading (Guskey, 1994). Others worried that measuring aptitude was different from measuring levels of classroom performance, which might not be normally distributed (Schneider and Hutt, 2013).

Based on the above research and the pressure toward uniformity of grading systems, by the 1940s the “A”–“F” grading system was dominant, with the four-point scale and percentages still also in use (Schneider and Hutt, 2013). However, many inconsistencies remained. As one example, Yale used no less than four different grading systems from the 1960s to 1980s (Yale University, 2013).

### *Present Day*

Grading systems remain controversial and hotly debated today (Jaschik, 2009). Some argue grades are psychologically harmful (Kohn, 1999). Others raise concerns about the integrity of the “A”–“F” system, given well-documented trends in grade inflation (Rojstaczer and Healy, 2012). One professor summed it up by saying grades do no more than “create a facade of coherence” (Jaschik, 2009). A number of colleges have abandoned numerical and categorical grading altogether, opting instead for creating contracts with students to define success or employing student self-reflection in combination with written evaluations by faculty (Jaschik, 2009). Among the Ivy League schools, Brown University does not calculate grade point averages, does not use “D’s” in its grading scale, and does not record failing grades (Brown University, 2014). Even Yale, the institution that started this history of grading more than 200 yr ago, is today still considering changes to its grading system (Yale University, 2013).

Though grades were initially meant to serve various pedagogical purposes, more recent reforms have focused on “grades as useful tools in an organizational rather than pedagogical enterprise—tools that would facilitate movement, communication, and coordination” (Schneider and Hutt, 2013). So, what are the potential purposes of grading in educational settings?

## **PURPOSES OF GRADING—PAST AND PRESENT**

### *Grades as Feedback on Performance—Does Grading Provide Feedback to Help Students Understand and Improve upon Their Deficiencies?*

[This] work affirms an observation that many classroom teachers have made about their students: if a paper is returned with both a grade and a comment, many students will pay attention to the grade and ignore the comment.

—Brookhart (2008, p. 8)

For most faculty members, the concept of feedback has at least two applications to the concept of grading. On one hand, grading itself is a form of feedback that may be

useful to students. In addition, in the process of grading student work, faculty members sometimes provide written comments as feedback that students could use to improve their work. Because college students express a desire for feedback (Higgins *et al.*, 2002), faculty members may feel pressured to grade more (rather than facilitating ungraded activities) and to provide more written feedback while grading. Especially in large classes, this can significantly increase workload on faculty (Nicol and Macfarlane-Dick, 2006; Crisp, 2007). But are grades and written comments effective forms of feedback that assist students in achieving conceptual mastery of the subject?

Feedback is generally divided into two categories: evaluative feedback and descriptive feedback. Evaluative feedback, such as a letter grade or written praise or criticism, judges student work, while descriptive feedback provides information about how a student can become more competent (Brookhart, 2008, p. 26). Butler and Nisan (1986) compared the impacts of evaluative feedback, descriptive feedback, and no feedback on student achievement in problem-solving tasks and in “quantitative” tasks (e.g., those requiring quick, timed work to produce a large number of answers). They found that students receiving descriptive feedback (but *not* grades) on an initial assignment performed significantly better on follow-up quantitative tasks and problem-solving tasks than did students receiving grades or students receiving no feedback. Students receiving grades performed better on follow-up quantitative tasks than students receiving no feedback, but did not outperform those students on problem-solving assignments. In other words, providing evaluative feedback (in this case, grades) after a task does not appear to enhance students’ future performance in problem solving.

While descriptive, written feedback can enhance student performance on problem-solving tasks; reaping those benefits requires students to read, understand, and use the feedback. Anecdotal accounts, as well as some studies, indicate that many students do not read written feedback, much less use it to improve future work (MacDonald, 1991; Crisp, 2007). In one study, less than half of undergraduate medical students even chose to collect the feedback provided on their essays (Sinclair and Cleland, 2007). Other studies suggest that many students do read feedback and consider it carefully but the feedback is written in a way that students do not find useful in improving future work (Higgins *et al.*, 2002). Some studies have further investigated the relationships between grading and descriptive feedback by providing students with both written feedback and grades on assignments. In these cases, the addition of written comments consistently failed to enhance student performance on follow-up tasks (Marble *et al.*, 1978; Butler 1988; Pulfrey *et al.*, 2011). Brookhart (2008, p. 8) concludes, “the grade ‘trumps’ the comment” and “comments have the best chance of being *read* as descriptive if they are not accompanied by a grade.” Even when written feedback is read, there is widespread agreement that instructor feedback is very difficult for students to interpret and convert into improved future performance (Weaver, 2006).

Grading does not appear to provide effective feedback that constructively informs students’ future efforts. This is particularly true for tasks involving problem solving or creativity. Even when grading comes in the form of written comments, it is unclear whether students even read such comments, much less understand and act on them.

### ***Grades as a Motivator of Student Effort—Does Grading Motivate Students to Learn?***

Our results suggest...that the information routinely given in schools—that is, grades—may encourage an emphasis on quantitative aspects of learning, depress creativity, foster fear of failure, and undermine interest.

—Butler and Nisan (1986)

As described in the history of grading above, our current “A”–“F” grading system was not designed with the primary intent of motivating students. Rather, it stemmed from efforts to streamline communication between institutions and diminish the impacts of unreliable evaluation of students from teacher to teacher (Grant and Green, 2013). That is not to say, however, that grades do not have an impact on student motivation and effort. At some point, every instructor has likely experienced desperate petitions from students seeking more points—a behavior that seems to speak to an underlying motivation stimulated by the grading process.

It would not be surprising to most faculty members that, rather than stimulating an interest in learning, grades primarily enhance students’ motivation to avoid receiving bad grades (Butler and Nisan, 1986; Butler, 1988; Crooks, 1988; Pulfrey *et al.*, 2011). Grades appear to play on students’ fears of punishment or shame, or their desires to outcompete peers, as opposed to stimulating interest and enjoyment in learning tasks (Pulfrey *et al.*, 2011). Grades can dampen existing intrinsic motivation, give rise to extrinsic motivation, enhance fear of failure, reduce interest, decrease enjoyment in class work, increase anxiety, hamper performance on follow-up tasks, stimulate avoidance of challenging tasks, and heighten competitiveness (Harter, 1978; Butler and Nisan, 1986; Butler, 1988; Crooks, 1988; Pulfrey *et al.*, 2011). Even providing encouraging, written notes on graded work does not appear to reduce the negative impacts grading exerts on motivation (Butler, 1988). Rather than seeing low grades as an opportunity to improve themselves, students receiving low scores generally withdraw from class work (Butler, 1988; Guskey, 1994). While students often express a desire to be graded, surveys indicate they would prefer descriptive comments to grades as a form of feedback (Butler and Nisan, 1986).

High-achieving students on initial graded assignments appear somewhat sheltered from some of the negative impacts of grades, as they tend to maintain their interest in completing future assignments (presumably in anticipation of receiving additional good grades; Butler, 1988). Oettinger (2002) and Grant and Green (2013) looked specifically for positive impacts of grades as incentives for students on the threshold between grade categories in a class. They hypothesized that, for example, a student on the borderline between a “C” and a “D” in a class would be more motivated to study for a final exam than a student solidly in the middle of the “C” range. However, these studies found only minimal (Oettinger, 2002) or no (Grant and Green, 2013) evidence that grades motivated students to perform better on final exams under these conditions.

This is not to say that classroom evaluation is by definition harmful or a thing to avoid. Evaluation of students in the service of learning—generally including a mechanism for feedback without grade assignment—can serve to enhance learning and motivation (Butler and Nisan, 1986; Crooks,

1988; Kitchen *et al.*, 2006). Swinton (2010) additionally found that a grading system that explicitly rewarded effort in addition to rewarding knowledge stimulated student interest in improvement. This implies that balancing accuracy-based grading with providing meaningful feedback and awarding student effort could help avoid some of the negative consequences of grading.

Rather than motivating students to learn, grading appears to, in many ways, have quite the opposite effect. Perhaps at best, grading motivates high-achieving students to continue getting high grades—regardless of whether that goal also happens to overlap with learning. At worst, grading lowers interest in learning and enhances anxiety and extrinsic motivation, especially among those students who are struggling.

### ***Grades as a Tool for Comparing Students—Is Grading on a Curve the Fairest Way to Grade?***

You definitely compete for grades in engineering; whereas you earn grades in other disciplines . . . I have to get one point higher on the test than the next guy so I can get the higher grade.

—Student quoted in Seymour and Hewitt (1997, p. 118)

The concept of grading on a curve arose from studies in the early 20th century suggesting that levels of aptitude, for example as measured by IQ, were distributed in the population according to a normal curve. Some then argued, if a classroom included a representative sample from the population, grades in the class should similarly be distributed according to a normal curve (Finkelstein, 1913). Conforming grades to a curve held the promise of addressing some of the problems surrounding grading by making the process more scientific and consistent across classrooms (Meyer, 1908). Immediately, even some proponents of curved grading recognized problems with comparing levels of aptitude in the population with levels of classroom achievement among a population of students. For a variety of reasons, a given classroom might not include a representative sample from the general population. In addition, teachers often grade based on a student's performance or accomplishment in the classroom—characteristics that differ in many ways from aptitude (Finkelstein, 1913). However, despite the reservations of some teachers and researchers, curved grading steadily gained acceptance throughout much of the 20th century (Schneider and Hutt, 2013).

Grading on a curve is by definition a type of “norm-referenced” grading, meaning student work is graded based on comparisons with other students' work (Brookhart, 2004, p. 72). One issue surrounding norm-referenced grading is that it can dissociate grades from any meaning in terms of content knowledge and learning. Bloom (1968) pointed out that, in grading on a curve “it matters not that the failures of one year performed at about the same level as the C students of another year. Nor does it matter that the A students of one school do about as well as the F students of another school.” As this example demonstrates, under curved grading, grades might not communicate any information whatsoever regarding a student's mastery of course knowledge or skills.

Of even more concern, however, is the impact norm-referenced grading has on competition between students. The quote at the start of this section describes how many stu-

dents respond to curve-graded classes compared with classes that do not use a grading curve. Seymour and Hewitt (1997, p. 118) explain, “Curve-grading forces students to compete with each other, whether they want to or not, because it exaggerates very fine degrees of differences in performance. Where there is little or no difference in work standards, it encourages a struggle to create it.” Studies have shown that science students in competitive class environments do not learn or retain information as well as students in cooperative class environments (Humphreys *et al.*, 1982). Students in cooperative environments are additionally more interested in learning and find learning more worthwhile than students in competitive environments (Humphreys *et al.*, 1982). Of particular concern is that the competitive environment fostered by norm-referenced grading represents one of the factors contributing to the loss of qualified, talented, and often underrepresented college students from science fields (Seymour and Hewitt, 1997; Tobias, 1990). Disturbingly, even when a science instructor does not grade on a curve, students might, due to their past experiences, assume a curve is used and adopt a competitive stance anyway (Tobias, 1990, p. 23).

Bloom (1968, 1976) presents evidence and a theoretical framework supporting an alternate view of grading whereby most students would be expected to excel and not fall into the middle grades. He states, “If the students are normally distributed with respect to aptitude, but the kind and quality of instruction and the amount of time available for learning are made appropriate to the characteristics and needs of each student, the majority of students may be expected to achieve mastery of the subject. And, the relationship between aptitude and achievement should approach zero” (Bloom, 1968). In other words, even if we were to accept a concept of innate aptitude that is normally distributed in a classroom, that distribution should not predict classroom achievement, provided the class environment supports diverse learners in appropriate ways. This idea was a significant development, because it freed teachers from the stigma associated with awarding a larger number of high grades. Previously, an excess of higher grades was thought to arise only from either cheating by students or poor grading practices by teachers (Meyer, 1908). Bloom's model argues that, when given the proper learning environment and compared against standards of mastery in a field (rather than against one another), large numbers of students could succeed. This type of grading—where instructional goals form the basis of comparison—is called “criterion-referenced” grading (Brookhart, 2004, p. 72).

Of course, Bloom's work did not rule out the possibility that some teachers might still give high grades for undesirable reasons unrelated to standards of mastery (e.g., to be nice, to gain the admiration of students, etc.). Such practices would not be in line with Bloom's work and would lead to pernicious grade inflation. Indeed, many of those bemoaning recent trends in grade inflation in higher education (though less prevalent in the sciences) point to the abandonment of curved grading as a major factor (Rojstaczer and Healy, 2012). Such studies often promote various forms of curving—at the level of individual courses or even at the institution as a whole—to combat inflation (Johnson, 2003, chaps. 7–8). In light of the above, however, it seems strange to aspire to introduce grading systems that could further push students into competition and give rise to grades that indicate little about the mastery of knowledge or skills in a subject. The broader

distribution of grades under curve-adjusted grading could simply create the illusion of legitimacy in the grading system without any direct connection between grades and achievement of learning goals. Perhaps the more productive route is to push for stronger, criterion-referenced grading systems in which instructional goals, assessments, and course work are more intimately aligned.

In brief, curved grading creates a competitive classroom environment, alienates certain groups of talented students, and often results in grades unrelated to content mastery. Curving is therefore not the fairest way to assign grades.

### ***Grades as an Objective Evaluation of Student Knowledge—Do Grades Provide Reliable Information about Student Learning?***

Study Critiques Schools over Subjective Grading: An Education Expert Calls for Greater Consistency in Evaluating Students' Work.

—Los Angeles Times (2009)

As evidenced by the above headline, some have criticized grading as subjective and inconsistent, meaning that the same student could receive drastically different grades for the same work, depending on who is grading the work and when it is graded. The literature indeed indicates that some forms of assessment lend themselves to greater levels of grading subjectivity than others.

Scoring multiple-choice assessments does not generally require the use of professional judgment from one paper to the next, so instructors should be able to score such assessments objectively (Wainer and Thissen, 1993; Anderson, 2008, p. 451). However, despite their advantages in terms of objective grading, studies have raised concerns regarding the blanket use of multiple-choice assessments. Problems with such assessments range from their potential to falsely indicate student understanding to the possibilities that they hamper critical thinking and exhibit bias against certain groups of students (Townes and Robinson, 1993; Scouller, 1998; Rogers and Harley, 1999; Paxton, 2000; Dufresne *et al.*, 2002; Zimmerman and Williams, 2003; Stanger-Hall, 2012).

Grading student writing, whether in essays, reports, or constructed-response test items, opens up greater opportunities for subjectivity. Shortly after the rise in popularity of percentage-based grading systems in the early 1900s, researchers began examining teacher consistency in marking written work by students. Starch and Elliott (1912) asked 142 teachers to grade the same English paper and found that grades on the paper varied from 50 to 98% between teachers. Because different teachers awarded scores ranging from failing to exceptional, the researchers concluded “the promotion or retardation of a pupil depends to a considerable extent upon the subjective estimate of his teacher” rather than upon the actual work produced by the student (Starch and Elliott, 1912). Even greater levels of inconsistency were found in teachers' scoring of a geometry paper showing the solution to a problem (Starch and Elliott, 1913).

Eells (1930) investigated the consistency of individual teachers' grading by asking 61 teachers to grade the same history and geography papers twice—the second time 11 wk after the first. He concluded that “variability of grading is about as great in the same individual as in groups of different

individuals” and that, after analysis of reliability coefficients, assignment of scores amounted to “little better than sheer guesses” (Eells, 1930). Similar problems in marking reliability have been observed in higher education environments, although the degree of reliability varies dramatically, likely due to differences in instructor training, assessment type, grading system, and specific topic assessed (Meadows and Billington, 2005, pp. 18–20). Factors that occasionally influence an instructor's scoring of written work include the penmanship of the author (Bull and Stevens, 1979), sex of the author (Spear, 1984), ethnicity of the author (Fajardo, 1985), level of experience of the instructor (Weigle, 1999), order in which the papers are reviewed (Farrell and Gilbert, 1960; Spear, 1996), and even the attractiveness of the author (Bull and Stevens, 1979).

Designing and using rubrics to grade assignments or tests can reduce inconsistencies and make grading written work more objective. Sharing the rubrics with students can have the added benefit of enhancing learning by allowing for feedback and self-assessment (Jonsson and Svingby, 2007; Reddy and Andrade, 2010). Consistency in grading tests can also be improved by writing longer tests with more narrowly focused questions, but this would tend to limit the types of questions that could appear on an exam (Meadows and Billington, 2005).

In summary, grades often fail to provide reliable information about student learning. Grades awarded can be inconsistent both for a single instructor and among different instructors for reasons that have little to do with a students' content knowledge or learning advances. Even multiple-choice tests, which can be graded with great consistency, have the potential to provide misleading information on student knowledge.

student learning by grading less or at least differently than they have before.

### **IN CONCLUSION—TEACHING MORE BY GRADING LESS (OR DIFFERENTLY)**

A review of the history and research on grading practices may appear to present a bleak outlook on the process of grading and its impacts on learning. However, underlying the less encouraging news about grades are numerous opportunities for faculty members to make assessment and evaluation more productive, better aligned with student learning, and less burdensome for faculty and students. Notably, many of the practices advocated in the literature would appear to involve faculty members spending less time grading. The time and energy spent on grading has been often pinpointed as a key barrier to instructors becoming more innovative in their teaching. In some cases, the demands of grading require so much instructor attention, little time remains for reflection on the structure of a course or for aspirations of pedagogical improvement. Additionally, some instructors are hesitant to develop active-learning activities—as either in-class activities or homework assignments—for fear of the onslaught of grading resulting from these new activities. However, just because students generate work does not mean instructors need to grade that work for accuracy. In fact, we have presented evidence that accuracy-based grading may, in fact, demotivate students and impede learning. Additionally, the time-consuming process of instructors marking papers and leaving comments may achieve no gain, if comments are rarely read by students. One wonders how much more student learning might occur if instructors' time spent grading was used in different ways. What if instructors spent more time planning in-class discussions of homework and simply assigned a small number of earned points to students for completing the work? What if students themselves used rubrics to examine their peers' efforts and evaluate their own work, instead of instructors spending hours and hours commenting on papers? What if students viewed their peers as resources and collaborators, as opposed to competitors in courses that employ grade curving? Implementing small changes like those described above might allow instructors to promote more