

The Case Against Grades

By Alfie Kohn

[This is a slightly expanded version of the published article.]

“I remember the first time that a grading rubric was attached to a piece of my writing....Suddenly all the joy was taken away. I was writing for a grade — I was no longer exploring for me. I want to get that back. Will I ever get that back?”

— Claire, a student (in Olson, 2006)

By now enough has been written about academic assessment to fill a library, but when you stop to think about it, the whole enterprise really amounts to a straightforward two-step dance. We need to collect information about how students are doing, and then we need to share that information (along with our judgments, perhaps) with the students and their parents. Gather and report — that’s pretty much it.

You say the devil is in the details? Maybe so, but I’d argue that too much attention to the particulars of implementation may be distracting us from the bigger picture — or at least from a pair of remarkable conclusions that emerge from the best theory, practice, and research on the subject: *Collecting information doesn’t require tests, and sharing that information doesn’t require grades.* In fact, students would be a lot better off without either of these relics from a less enlightened age.

Why tests are not a particularly useful way to assess student learning (at least the kind that matters), and what thoughtful educators do instead, are questions that must wait for **another day**. Here, our task is to take a hard look at the second practice, the use of letters or numbers as evaluative summaries of how well students have done, regardless of the method used to arrive at those judgments.

The Effects of Grading

Most of the criticisms of grading you’ll hear today were laid out forcefully and eloquently anywhere from four to eight decades ago (Crooks, 1933; De Zouche, 1945; Kirschenbaum, Simon, & Napier, 1971; Linder, 1940; Marshall, 1968), and these early essays make for eye-opening reading. They remind us just how long it’s been clear there’s something wrong with what we’re doing as well as just how little progress we’ve made in acting on that realization.

In the 1980s and '90s, educational psychologists systematically studied the effects of grades. As I've reported elsewhere (Kohn, 1999a, 1999b, 1999c), when students from elementary school to college who are led to focus on grades are compared with those who aren't, the results support three robust conclusions:

* *Grades tend to diminish students' interest in whatever they're learning.* A "grading orientation" and a "learning orientation" have been shown to be inversely related and, as far as I can tell, every study that has ever investigated the impact on intrinsic motivation of receiving grades (or instructions that emphasize the importance of getting good grades) has found a negative effect.

* *Grades create a preference for the easiest possible task.* Impress upon students that what they're doing will count toward their grade, and their response will likely be to avoid taking any unnecessary intellectual risks. They'll choose a shorter book, or a project on a familiar topic, in order to minimize the chance of doing poorly — not because they're "unmotivated" but because they're rational. They're responding to adults who, by telling them the goal is to get a good mark, have sent the message that success matters more than learning.

* *Grades tend to reduce the quality of students' thinking.* They may skim books for what they'll "need to know." They're less likely to wonder, say, "How can we be sure that's true?" than to ask "Is this going to be on the test?" In one experiment, students told they'd be graded on how well they learned a social studies lesson had more trouble understanding the main point of the text than did students who were told that no grades would be involved. Even on a measure of rote recall, the graded group remembered fewer facts a week later (Grolnick and Ryan, 1987).

Research on the effects of grading has slowed down in the last couple of decades, but the studies that are still being done reinforce the earlier findings. For example, a grade-oriented environment is associated with increased levels of cheating (Anderman and Murdock, 2007), grades (whether or not accompanied by comments) promote a fear of failure even in high-achieving students (Pulfrey et al., 2011), and the elimination of grades (in favor of a pass/fail system) produces substantial benefits with no apparent disadvantages in medical school (White and Fantone, 2010). More important, no recent research has contradicted the earlier "big three" findings, so those conclusions still stand.

Why Grading Is Inherently Problematic

A student asked his Zen master how long it would take to reach enlightenment. "Ten years," the master said. But, the student persisted, what if he studied very hard? "Then 20 years," the master responded. Surprised, the student asked how long it would take if he worked very, very hard and became the most dedicated student in the Ashram. "In that case, 30

years,” the master replied. His explanation: “If you have one eye on how close you are to achieving your goal, that leaves only one eye for your task.”

To understand why research finds what it does about grades, we need to shift our focus from educational measurement techniques to broader psychological and pedagogical questions. The latter serve to illuminate a series of misconceived assumptions that underlie the use of grading.

Motivation: While it’s true that many students, after a few years of traditional schooling, could be described as motivated by grades, what counts is the nature of their motivation. Extrinsic motivation, which includes a desire to get better grades, is not only different from, but often undermines, intrinsic motivation, a desire to learn for its own sake (Kohn 1999a). Many assessment specialists talk about motivation as though it were a single entity — and their recommended practices just put a finer gloss on a system of rewards and punishments that leads students to chase marks and become less interested in the learning itself. If nourishing their *desire* to learn is a primary goal for us, then grading is problematic by its very nature.

Achievement: Two educational psychologists pointed out that “an overemphasis on assessment can actually undermine the pursuit of excellence” (Maehr and Midgley, 1996, p. 7). That unsettling conclusion — which holds regardless of the quality of the assessment but is particularly applicable to the use of grades — is based on these researchers’ own empirical findings as well as those of many others, including Carol Dweck, Carole Ames, Ruth Butler, and John Nicholls (for a review, see Kohn 1999b, chapter 2). In brief: the more students are led to focus on *how well* they’re doing, the less engaged they tend to be with *what* they’re doing.

It follows that all assessment must be done carefully and sparingly lest students become so concerned about their achievement (how good they are at doing something — or, worse, how their performance compares to others’) that they’re no longer thinking about the learning itself. Even a well-meaning teacher may produce a roomful of children who are so busy monitoring their own reading skills that they’re no longer excited by the stories they’re reading. Assessment consultants worry that grades may not accurately reflect student performance; educational psychologists worry because grades fix students’ attention *on* their performance.

Quantification: When people ask me, a bit defensively, if it isn’t important to measure how well students are learning (or teachers are teaching), I invite them to rethink their choice of verb. There is certainly value in *assessing* the quality of learning and teaching, but that doesn’t mean it’s always necessary, or even possible, to *measure* those things — that is, to turn them into numbers. Indeed, “measurable outcomes may be the least significant results

of learning” (McNeil, 1986, p. xviii) — a realization that offers a refreshing counterpoint to today’s corporate-style “school reform” and its preoccupation with data.

To talk about what happens in classrooms, let alone in children’s heads, as moving forward or backward in specifiable degrees, is not only simplistic because it fails to capture much of what is going on, but also destructive because it may change what is going on for the worse. Once we’re compelled to focus only on what can be reduced to numbers, such as how many grammatical errors are present in a composition or how many mathematical algorithms have been committed to memory, thinking has been severely compromised. And that is exactly what happens when we try to fit learning into a four- or five- or (heaven help us) 100-point scale.

Curriculum: “One can have the best assessment imaginable,” Howard Gardner (1991, p. 254) observed, “but unless the accompanying curriculum is of quality, the assessment has no use.” Some people in the field are candid about their relativism, offering to help align your assessment to whatever your goals or curriculum may be. The result is that teachers may become more adept at measuring how well students have mastered a collection of facts and skills whose value is questionable — and never questioned. “If it’s not worth teaching, it’s not worth teaching well,” as Eliot Eisner (2001, p. 370) likes to say. Nor, we might add, is it worth assessing accurately.

Portfolios, for example, can be constructive if they replace grades rather than being used to *yield* them. They offer a way to thoughtfully gather a variety of meaningful examples of learning for the students to review. But what’s the point, “if instruction is dominated by worksheets so that every portfolio looks the same”? (Neill et al. 1995, p. 4). Conversely, one sometimes finds a mismatch between more thoughtful forms of pedagogy — say, a workshop approach to teaching writing — and a depressingly standardized assessment tool like rubrics (Wilson, 2006).

Improving Grading: A Fool’s Errand?

“I had been advocating standards-based grading, which is a very important movement in its own right, but it took a push from some great educators to make me realize that if I wanted to focus my assessment around authentic feedback, then I should just abandon grades altogether.”

— New Jersey middle school teacher Jason Bedell (2010)

Much of what is prescribed in the name of “assessing for learning” (and, for that matter, “formative assessment”) leaves me uneasy: The recommended practices often seem prefabricated and mechanistic; the imperatives of data collection seem to upstage the

children themselves and the goal of helping them become more enthusiastic about what they're doing. Still, if it's done only occasionally and with humility, I think it's possible to assess for learning. But *grading* for learning is, to paraphrase a 1960's-era slogan, rather like bombing for peace. Rating and ranking students (and their efforts to figure things out) is inherently counterproductive.

If I'm right — more to the point, if all the research to which I've referred is taken seriously — then the absence of grades is a necessary, though not sufficient, condition for promoting deep thinking and a desire to engage in it. It's worth lingering on this proposition in light of a variety of efforts to sell us formulas to improve our grading techniques, none of which address the problems of grading, per se.

* It's not enough to replace letters or numbers with labels ("exceeds expectations," "meets expectations," and so on). If you're sorting students into four or five piles, you're still grading them. Rubrics typically include numbers as well as labels, which is only one of several reasons they merit our skepticism (Wilson, 2006; Kohn, 2006).

* It's not enough to tell students in advance exactly what's expected of them. "When school is seen as a test, rather than an adventure in ideas," teachers may persuade themselves they're being fair "if they specify, in listlike fashion, exactly what must be learned to gain a satisfactory grade...[but] such schooling is unfair in the wider sense that it prepares students to pass other people's tests without strengthening their capacity to set their own assignments in collaboration with their fellows" (Nicholls and Hazzard, 1993, p. 77).

* It's not enough to disseminate grades more efficiently — for example, by posting them on-line. There is a growing technology, as the late Gerald Bracey once remarked, "that permits us to do in nanoseconds things that we shouldn't be doing at all" (quoted in Mathews, 2006). In fact, posting grades on-line is a significant step backward because it enhances the salience of those grades and therefore their destructive effects on learning.

* It's not enough to add narrative reports. "When comments and grades coexist, the comments are written to justify the grade" (Wilson, 2009, p. 60). Teachers report that students, for their part, often just turn to the grade and ignore the comment, but "when there's only a comment, they read it," says high school English teacher Jim Drier. Moreover, research suggests that the harmful impact of grades on creativity is no less (and possibly even more) potent when a narrative accompanies them. Narratives are helpful only in the absence of grades (Butler, 1988; Pulfrey et al., 2011).

* It's not enough to use "standards-based" grading. That phrase may suggest any number of things — for example, more consistency, or a reliance on more elaborate formulas, in determining grades; greater specificity about what each grade signifies; or an increase in the

number of tasks or skills that are graded. At best, these prescriptions do nothing to address the fundamental problems with grading. At worst, they exacerbate those problems. In addition to the simplistic premise that it's always good to have more data, we find a penchant shared by the behaviorists of yesteryear that learning can and should be broken down into its components, each to be evaluated separately. And more frequent temperature-taking produces exactly the kind of disproportionate attention to performance (at the expense of learning) that researchers have found to be so counterproductive.

The term “standards-based” is sometimes intended just to mean that grading is aligned with a given set of objectives, in which case our first response should be to inquire into the value of those objectives (as well as the extent to which students were invited to help formulate them). If grades are based on state standards, there's particular reason to be concerned since those standards are often too specific, age-inappropriate, superficial, and standardized by definition. In my experience, the best teachers tend to be skeptical about aligning their teaching to a list imposed by distant authorities, or using that list as a basis for assessing how well their students are thinking.

Finally, “standards-based” may refer to something similar to criterion-based testing, where the idea is to avoid grading students on a curve. (Even some teachers who don't do so explicitly nevertheless act as though grades ought to fall into something close to a normal distribution, with only a few students receiving As. But this pattern is not a fact of life, nor is it a sign of admirable “rigor” on the teacher's part. Rather, “it is a symbol of failure — failure to teach well, failure to test well, and failure to have any influence at all on the intellectual lives of students” [Milton, Pollio, & Eison, 1986].) This surely represents an improvement over a system in which the number of top marks is made artificially scarce and students are set against one another. But here we've peeled back the outer skin of the onion (competition) only to reveal more noxious layers beneath: extrinsic motivation, numerical ratings, the tendency to promote achievement at the expense of learning.

If we begin with a desire to assess more often, or to produce more data, or to improve the consistency of our grading, then certain prescriptions will follow. If, however, our point of departure isn't mostly about the grading, but about our desire for students to understand ideas from the inside out, or to get a kick out of playing with words and numbers, or to be in charge of their own learning, then we will likely end up elsewhere. We may come to see grading as a huge, noisy, fuel-guzzling, smoke-belching machine that constantly requires repairs and new parts, when what we should be doing is pulling the plug.

Deleting — or at Least Diluting — Grades

“Like it or not, grading is here to stay” is a statement no responsible educator would ever offer as an excuse for inaction. What matters is whether a given practice is in the best interest of

students. If it isn't, then our obligation is to work for its elimination and, in the meantime, do what we can to minimize its impact.

Replacing letter and number grades with narrative assessments or conferences — qualitative summaries of student progress offered in writing or as part of a conversation — is not a utopian fantasy. It has already been done successfully in many elementary and middle schools and even in some high schools, both public and private (Kohn, 1999c). It's important not only to realize that such schools exist but to investigate *why* they've eliminated grades, how they've managed to do so (hint: the process can be gradual), and what benefits they have realized.

Naturally objections will be raised to this — or any — significant policy change, but once students and their parents have been shown the relevant research, reassured about their concerns, and invited to participate in constructing alternative forms of assessment, the abolition of grades proves to be not only realistic but an enormous improvement over the status quo. Sometimes it's only after grading has ended that we realize just how harmful it's been.

To address one common fear, the graduates of grade-free high schools are indeed accepted by selective private colleges and large public universities — on the basis of narrative reports and detailed descriptions of the curriculum (as well as recommendations, essays, and interviews), which collectively offer a fuller picture of the applicant than does a grade-point average. Moreover, these schools point out that their students are often more motivated and proficient learners, thus better prepared for college, than their counterparts at traditional schools who have been preoccupied with grades.

In any case, college admission is surely no bar to eliminating grades in elementary and middle schools because colleges are largely indifferent to what students have done before high school. That leaves proponents of grades for younger children to fall back on some version of an argument I call "BGUTI": Better Get Used To It (Kohn, 2005). The claim here is that we should do unpleasant and unnecessary things to children now in order to prepare them for the fact that just such things will be done to them later. This justification is exactly as absurd as it sounds, yet it continues to drive education policy.

Even when administrators aren't ready to abandon traditional report cards, individual teachers can help to rescue learning in their own classrooms with a two-pronged strategy to "neuter grades," as one teacher described it. First, they can stop putting letter or number grades on individual assignments and instead offer only qualitative feedback. Report cards are bad enough, but the destructive effects reported by researchers (on interest in learning, preference for challenge, and quality of thinking) are compounded when students are rated on what they do in school day after day. Teachers can mitigate considerable harm by

replacing grades with authentic assessments; moreover, as we've seen, any feedback they may already offer becomes much more useful in the absence of letter or number ratings.

Second, although teachers may be required to submit a final grade, there's no requirement for them to decide unilaterally what that grade will be. Thus, students can be invited to participate in that process either as a negotiation (such that the teacher has the final say) or by simply permitting students to grade themselves. If people find that idea alarming, it's probably because they realize it creates a more democratic classroom, one in which teachers must create a pedagogy and a curriculum that will truly engage students rather than allow teachers to coerce them into doing whatever they're told. In fact, negative reactions to this proposal ("It's unrealistic!") point up how grades function as a mechanism for controlling students rather than as a necessary or constructive way to report information about their performance.

I spoke recently to several middle and high school teachers who have de-graded their classes. Jeff Robbins, who has taught eighth-grade science in New Jersey for 15 years, concedes that "life was easier with grades" because they take so much less time than meaningful assessment. That efficiency came at a huge cost, though, he noticed: Kids were stressed out and also preferred to avoid intellectual risks. "They'll take an easier assignment that will guarantee the A."

Initially Robbins announced that any project or test could be improved and resubmitted for a higher grade. Unfortunately, that failed to address the underlying problem, and he eventually realized he had to stop grading entirely. Now, he offers comments to all of his 125 students "about what they're doing and what they need to improve on" and makes abbreviated notes in his grade book. At the end of the term, over a period of about a week, he grabs each student for a conversation at some point — "because the system isn't designed to allow kids this kind of feedback" — asking "what did you learn, how did you learn it. Only at the very end of the conversation [do] I ask what grade will reflect it... and we'll collectively arrive at something." Like many other teachers I've spoken to over the years, Robbins says he almost always accepts students' suggestions because they typically pick the same grade that he would have.

Jim Drier, an English teacher at Mundelein High School in Illinois who has about 90 students ranging "from at-risk to A.P.," was relieved to find that it "really doesn't take that long" to write at least a brief note on students' assignments — "a reaction to what they did and some advice on how they might improve." But he never gives them "a number or grade on anything they do. The things that grades make kids do are heartbreaking for an educator": arguing with teachers, fighting with parents, cheating, memorizing facts just for a test and then forgetting them. "This is not why I became a teacher."

Without grades, “I think my relationships with students are better,” Drier says. “Their writing improves more quickly and the things they learn stay with them longer. I’ve had lots of kids tell me it’s changed their attitude about coming to school.” He expected resistance from parents but says that in three years only one parent has objected, and it may help that he sends a letter home to explain exactly what he’s doing and why. Now two of his colleagues are joining him in eliminating grades.

Drier’s final grades are based on students’ written self-assessments, which, in turn, are based on their review of items in their portfolios. He meets with about three-quarters of them twice a term, in most cases briefly, to assess their performance and, if necessary (although it rarely happens) to discuss a concern about the grade they’ve suggested. Asked how he manages without a grade book full of letters or numbers, Drier replies, “If I spend 18 weeks with them, I have a pretty good idea what their writing and reasoning ability is.”

A key element of authentic assessment for these and other teachers is the opportunity for students to help design the assessment and reflect on its purposes — individually and as a class. Notice how different this is from the more common variant of self-assessment in which students merely monitor their progress toward the teacher’s (or legislature’s) goals and in which they must reduce their learning to numerical ratings with grade-like rubrics.

Points of overlap as well as divergence emerge from the testimonies of such teachers, some of which have been collected by Joe Bower (n.d.), an educator in Red Deer, Alberta. Some teachers, for example, *evaluate* their students’ performance (in qualitative terms, of course), but others believe it’s more constructive to offer only *feedback* — which is to say, information. On the latter view, “the alternative to grades is description” and “the starting point for description is a plain sheet of paper, not a form which leads and homogenizes description” (Marshall, 1968, pp. 131, 143).

Teachers also report a variety of reactions to de-grading not only from colleagues and administrators but also from the students themselves. John Spencer (2010), an Arizona middle school teacher, concedes that “many of the ‘high performing’ students were angry at first. They saw it as unfair. They viewed school as work and their peers as competitors....Yet, over time they switch and they calm down. They end up learning more once they aren’t feeling the pressure” from grades.

Indeed, research suggests that the common tendency of students to focus on grades doesn’t reflect an innate predilection or a “learning style” to be accommodated; rather, it’s due to having been led for years to work for grades. In one study (Butler, 1992), some students were encouraged to think about how well they performed at a creative task while others were just invited to be imaginative. Each student was then taken to a room that contained a pile of pictures that other people had drawn in response to the same instructions. It also contained

some information that told them how to figure out their “creativity score.” Sure enough, the children who were told to think about their performance now wanted to know how they had done relative to their peers; those who had been allowed to become immersed in the task were more interested in seeing *what* their peers had done.

Grades don’t prepare children for the “real world” — unless one has in mind a world where interest in learning and quality of thinking are unimportant. Nor are grades a necessary part of schooling, any more than paddling or taking extended dictation could be described that way. Still, it takes courage to do right by kids in an era when the quantitative matters more than the qualitative, when meeting (someone else’s) standards counts for more than exploring ideas, and when anything “rigorous” is automatically assumed to be valuable. We have to be willing to challenge the conventional wisdom, which in this case means asking not how to improve grades but how to jettison them once and for all.

Why Grades Are Not Paramount to Achievement

The intrinsic love of learning supplants the drive for high marks in the long run.

Ashley Lamb-Sinclair 6/16/17 (The Atlantic)

At the beginning of this school year, my colleagues and I decided to avoid giving the sophomores in our English classes any grades for six weeks. Research [shows](#) that providing students with a number or letter in addition to quality comments prevents them from authentically reflecting. Quantitative grades also [diminish](#) student interest in learning, reduce academic risk taking, and decrease the quality of thinking. But beyond academics, as teachers, we saw the negative impact grades made on our students’ mental and emotional health. In fact, though a bit outdated, a 2002 study conducted by a psychologist at the University of Michigan showed 80 percent of students based their self worth on their academic success, [leading to](#) low self-esteem and other mental-health issues. In a highly academic setting, here was an opportunity to catalyze our students’ broader motivations for learning—a quality with macro-benefits in an environment obsessed with single Scantron marks.

Ours is the type of school that when I input grades into our electronic gradebook system, it is usually a matter of minutes before students knock on my classroom door and parents send me emails questioning single points and marks. It is frustrating for me to have the work I do

distilled down to a data point, so I was excited to try something new without the pressures of the gradebook. Our principal approved the plan and we sent a letter to families explaining the research and our reasoning for withholding grades. I also set up a closed Facebook group where I shared daily images, and I instituted weekly emails to keep parents abreast of the learning process.

All was mostly well during those six weeks, and I found my students to be incredibly engaged. At one point, we spent three days on a single thesis statement alone—students writing their claims, discussing them without fear of negative retribution in the form of a grade, and then redoing them happily. Before, such a lengthy lesson would have been met with endless questions about how many points it was worth, and if the teenage math didn't add up for them—if the cost outweighed the benefits—I would have been forced to increase the points to improve motivation or move on more quickly than they needed. In this situation, however, everything was low risk for the students, so they approached assignments more positively. And because grades were irrelevant, the challenge of the assignment took precedence.

If a student wasn't meeting expectations or turning in assignments, I discussed my concerns with them and their parents within a day. Many parents seemed to appreciate this quick, frequent contact, and several contacted me regularly for updates. I wasn't overwhelmed with this level of contact, although I might have been in the past, because I didn't have to worry about converting standards to point values and inputting them into the grade book. Typically, my assessment process takes anywhere from a week to a couple of months, and the pressure along the way from students, parents, and administrators often shifts my focus away from learning and toward frequent inputting of grades. But when grades were put on the back burner, I was freed up to focus on authentic communication of actual learning and growth. My sole focus became the learning of each student and sharing my observations of that learning in the moment. It felt like a revelation.

But then the time came to actually start grading. Parents who had been thrilled with consistent contact and specific feedback only weeks before suddenly seemed perplexed as to why their child's grades were not 100 percent all the time. I was soon spending significant time on the phone, responding to emails, and sitting down to conference with parents who complained about a lack of communication. Ironically, my contact with parents was the most intensive it had ever been in 12 years of teaching.

Now, obviously, I am not perfect. With any of these interactions, I very possibly could have been more communicative. But it was so baffling to me to see that in the eyes of a parent, a numerical evaluation was more informative and meaningful than frequent written and spoken descriptions. As the dust settled and we moved further into the school year, it occurred to me that perhaps this was because the product of learning is often more comfortable and affirming than the process of it. Consider that having a degree is often

validating regardless of actual skills, and a test score could never illustrate the hours spent studying to achieve it. A willingness to learn for its own sake **represents** intrinsic motivation, while grades and other accolades represent extrinsic. Research **has shown** time and again that intrinsic motivation leads to more profound learning. The truth is that the willingness to learn leads to achievement, but so often achievement is the only part that matters to others.

It is difficult to ignore the realities of grades and their extrinsic power. As frustrated as I was with the parents who wanted conferences or who argued over a point or two, I understand their frustration. They wanted what was best for their child which includes a quality college education and well-paying career. Test scores and grades certainly matter in the current academic system for achieving these goals. But while some employers do **still look** for higher GPAs for recent college graduates, most are **more concerned** with relevant internship, job, and volunteer experience. And extrinsic motivation **is often** effective when it is immediate, so graduating from high school and college with higher grades actually **represents** intrinsic motivation and persistence. So while the grades and scores matter as a quick reference for admissions officers and employers, they actually represent something more profound in the type of learner and worker in which they plan to invest.

Many professionals, including teachers, seek achievements to prove their value just as students do. Receiving National Board certification is often considered the gold standard in achievement for teachers. I have rarely encountered a National Board-certified teacher who was not a quality educator, and principals are consistently impressed when they learn I have the accolade. The value of the certification, though, comes from the process of earning it, not the framed certificate on the wall. Becoming National Board-certified is intensive and challenging. The extrinsic motivation of higher pay, more opportunities, and elevated status might have initially led me to seek National Board certification, but my dedication to improving as a professional guided me through the sometimes grueling process. I remember watching a video of myself teaching, stopping it minute-by-minute, and recording what was going on in the classroom in that instance. In the day-to-day whirlwind of managing student behavior, lesson plans, and grades, I don't have the luxury of such reflection and analysis. When my students spent three days on a single thesis statement, they practiced similar intensive reflection to when I was working toward becoming National Board-certified. And for both my students and for me, such opportunities are rare. So while parents might see a 96 percent in the gradebook and feel comforted by such a number, many don't actually know the work that led to it. The problem lies when the product itself is elevated above the process, and questions of improving revolve around getting an A and not mastering skills.

Even for students who are highly motivated by grades and test scores, I have seen the ugly side of this behavior. But during the six-week period when my students were freed from the pressures of grades and instead focused on being learners, engagement, enthusiasm,

motivation, and determination drastically **increased**. If only I could have measured those skills and shown them to my students' parents in a way that resonated with them.

But one parent in particular did respond positively—my principal. His daughter was in my class. Near the end of the six-week period, he sent me an email and wrote in it,

“I love the fact that you were willing to try something that to the outside world looked odd and probably drew dozens of ‘That’s not the way it was done when I was in school’ types of comments at dinner tables over the last month or so...Thank you, Ashley, for taking risks and putting your students’ learning above everything else.”

I appreciated that he realized what I was trying to do, but more than that, I appreciated that he understood why. He knew this because he is a parent and an educator. As both myself, what I want for my children is a love of learning, a driving passion for being better at something that matters to them, and educators who know the power of such desire as well.

When I was in high school, I took AP Chemistry II, a class I had no business taking because I was deeply unqualified and too immature to do so. I didn't care about learning high-level chemistry, and I was not the kind of student who was ever motivated by grades. So pretty early into the school year, I knew I was going to fail. The teacher was challenging, but caring and compassionate. I showed up to his class every day, tried to distract my friends—who did care about chemistry and grades—and when that didn't work, proceeded to put my head down and sleep. I failed the class, but I learned a valuable lesson. I stopped pushing myself to do things that I didn't believe in, and I focused my energy instead on things I did. To this day, it is the only class I have ever failed, but I wear the grade as a badge of honor. That sense of authentic self-identity guided me through college, graduate school, student teaching, writing resumes, landing jobs I loved, and, yes, eventually achievements of which I am proud.

Teaching More by Grading Less (or Differently)

By Jeffrey Schinske and Kimberly Tanner

[CBE Life Sci Educ.](#) 2014 Summer; 13(2): 159–166.

Abstract

The authors explore a history of grading and review the literature regarding the purposes and impacts of grading. They then suggest strategies for making grading more supportive of learning, including balancing accuracy-based and effort-based grading, using self/peer evaluation, curtailing curved grading, and exercising skepticism about the meaning of grades.

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 *Perjores* (Stiles, 1901, cited by Smallwood, 1935). By 1837, Yale was also recording 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 students 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 (Townsend 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 student’s 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.

GRADING—STRATEGIES FOR CHANGE

In part, grading practices in higher education have been driven by educational goals such as providing feedback to students, motivating students, comparing students, and measuring learning. However, much of the research literature on grading reviewed above suggests that these goals are often not being achieved with our current grading practices. Additionally, the expectations, time, and stress associated with grading may be distracting instructors from integrating other pedagogical practices that could create a more positive and effective classroom environment for learning. Below we explore several changes in approaching grading that could assist instructors in minimizing its negative influences. Kitchen *et al.* (2006) additionally provide an example of a high-enrollment college biology class that was redesigned to “maximize feedback and minimize the impact of grades.”

Balancing Accuracy-Based Grading with Effort-Based Grading

Multiple research studies described above suggest that the evaluative aspect of grading may distract students from a focus on learning. While evaluation will no doubt always be key in determining course grades, the entirety of students’ grades need not be based primarily on work that rewards only correct answers, such as exams and quizzes. Importantly, constructing a grading system that rewards students for participation and effort has been shown to stimulate student interest in improvement (Swinton, 2010). One strategy for focusing students on the importance of effort and practice in learning is to provide students opportunities to earn credit in a course for simply doing the work, completing assigned tasks, and engaging with the material. Assessing effort and participation can happen in a variety of ways (Bean and Peterson, 1998 ; Rocca, 2010). In college biology courses, clicker questions graded on participation and not correctness of responses is one strategy. Additionally, instructors can have students turn in minute papers in response to a question posed in class and reward this effort based on submission and not scientific accuracy. Perhaps most importantly, biology instructors can assign out-of-class work—case studies, concept maps, and other written assignments—that can promote student practice and focus students’ attention on key ideas, while not creating more grading work for the instructor. Those out-of-class assignments can be graded quickly (and not for accuracy) based on a simple rubric that checks whether students turned the work in on time, wrote the required minimum number of words, posed the required number of questions, and/or included a prescribed number of references. In summary, one strategy for changing grading is to balance accuracy-based grading with the awarding of some proportion of the grade based on student effort and participation. Changing grading in this way has the potential to promote student practice, incentivize in-class participation, and avoid some of the documented negative consequences of grading.

Providing Opportunities for Meaningful Feedback through Self and Peer Evaluation

Instructors often perceive grading to be a separate process from teaching and learning, yet well-crafted opportunities for evaluation can be effective tools for changing students' ideas about biology. Nicol and Macfarlane-Dick (2006) argue that, just as teaching strategies are shifting away from an instructor-centered, transmissionist approach to a more collaborative approach between instructor and students, so too should classroom feedback and grading. Because feedback traditionally has been given by the instructor and transmitted to students, Nicol and Macfarlane-Dick argue that students have been deprived of opportunities to become self-regulated learners who can detect their own errors in thinking. They advocate for incorporating techniques such as self-reflection and student dialogue into the assessment process. This, they hypothesize, would create feedback that is relevant to and understood by students and would release faculty members from some of the burden of writing descriptive feedback on student submissions. Additionally, peer review and grading practices can be the basis of in-class active-learning exercises, guided by an instructor-developed rubric. For example, students may be assigned out of class homework to construct a diagram of the flow of a carbon atom from a dead body to a coyote (Ebert-May *et al.*, 2003). With the development of a simple rubric, students can self- or peer-evaluate these diagrams during the next class activity to check for the inclusion of key processes, as determined by the instructor. The use of in-class peer evaluation thus allows students to see other examples of biological thinking beyond their own and that of the instructor. In addition, self-evaluation of one's own work using the instructor's rubric can build metacognitive skills in assessing one's own confusions and making self-corrections. Such evaluations need not take much time, and they have the potential to provide feedback that is meaningful and integrated into the learning process. In summary, both self- and peer-evaluation of work are avenues for providing meaningful feedback without formal grading on correctness that can positively influence students' learning (Sadler and Good, 2006 ; Freeman *et al.*, 2007 ; Freeman and Parks, 2010).

Making the Move Away from Curving

As documented in the research literature, the practice of grade curving has had unfortunate and often unintended consequences for the culture of undergraduate science classrooms, pitting students against one another as opposed to creating a collaborative learning community (Tobias, 1990 ; Seymour and Hewitt, 1997). As such, one simple adjustment to grading would be to abandon grading on a curve. Because the practice of curving is often assumed by students to be practiced in science courses, a move away from curving would likely necessitate explicit and repeated communication with students to convey that they are competing only against themselves and not one another. Moving away from curving sets the expectation that all students have the opportunity to achieve the highest possible grade. Perhaps most importantly, a move away from curving practices in grading may remove a key

remaining impediment to building a learning community in which students are expected to rely on and support one another in the learning process. In some instances, instructors may feel the need to use a curve when a large proportion of students perform poorly on a quiz or exam. However, an alternative approach would be to identify why students performed poorly and address this more specifically. For example, if the wording of an exam question was confusing for large numbers of students, then curving would not seem to be an appropriate response. Rather, excluding that question from analysis and in computing the exam grade would appear to be a more fair approach than curving. Additionally, if large numbers of students performed poorly on particular exam questions, providing opportunities for students to revisit, revise, and resubmit those answers for some credit would likely achieve the goal of not having large numbers of students fail. This would maintain the criterion-referenced grading system and additionally promote learning of the material that was not originally mastered. In summary, abandoning curving practices in undergraduate biology courses and explicitly conveying this to students could promote greater classroom community and student collaboration, while reducing well-documented negative consequences of this grading practice (Humphreys *et al.*, 1982).

Becoming Skeptical about What Grades Mean

The research literature raises significant questions about what grades really measure. However, it is likely that grades will continue to be the currency of formal teaching and learning in most higher education settings for the near future. As such, perhaps the most important consideration for instructors about grading is to simply be skeptical about what grades mean. Some instructors will refuse to write letters of recommendation for students who have not achieved grades in a particular range in their course. Yet, if grades are not a reliable reflection of learning and reflect other factors—including language proficiency, cultural background, or skills in test taking—this would seem a deeply biased practice. One practical strategy for making grading more equitable is to grade student work anonymously when possible, just as one would score assays in the laboratory blind to the treatment of the sample. The use of rubrics can also help remove bias from grading (Allen and Tanner, 2006) by increasing grading consistency. Perhaps most importantly, sharing grading rubrics with students can support them in identifying where their thinking has gone wrong and promote learning (Jonsson and Svingby, 2007 ; Reddy and Andrade, 2010). Much is yet to be understood about what influences students' performance in the context of formal education, and some have suggested grades may be more of a reflection of a students' ability to understand and play the game of school than anything to do with learning (Towns and Robinson, 1993 ; Scouller, 1998 ; Stanger-Hall, 2012). In summary, using tools such as rubrics and blind scoring in grading can decrease the variability and bias in grading student work. Additionally, remembering that grades are likely an inaccurate reflection of student learning can decrease assumptions instructors make about students.

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