EDUCATIONAL LEADERSHIP

| Home | Current Issue | Archives | Buy | Contact |

Reporting Student Learning

Ken O'Connor and Rick Wormeli

Despite advances in grading and reporting, imprecision and lack of meaning persist.

Back in 2000, Robert Marzano pointed out the rationale for changing grading practices. "Grades," he wrote, "are so imprecise that they are almost meaningless" (p. 1). Eleven years later, despite advances in grading and reporting in many schools and districts, this imprecision and lack of meaning persist.

It's time to evolve our grading practices. We believe there are four primary characteristics of effective grading. Grades should be accurate, consistent, meaningful, and supportive of learning. Let's examine why these characteristics are so important and how we can achieve them.

BUY THIS ISSUE

Read Abstract

Accurate

The Problem with Including Nonacademic Factors

Including whether a student maintained an organized notebook in his geometry grade dilutes the report of that student's geometry learning. "Organized notebook" is not a geometry standard. It's a helpful learning tool, of course, and wise teachers encourage students to take high-quality notes. However, we grade against standards and learner outcomes—not against the methods students use to achieve them. Instructional decisions made on the basis of these "fudged" grade reports are suspect; the reports offer no precise documentation and render descriptive feedback impossible.

The Problem with Grading Group Work

Suppose students work collaboratively in a history class to analyze rhetoric, prepare for debates, or prepare a multi-media presentation that analyzes economic models. These are all methods for teaching students the history curriculum, but they are not the history curriculum itself. In addition, when students

present their final report with everyone's names displayed on the opening slide, we're not sure where one student's influence ends and another's begins: To what extent does J. J. know the information without assistance from Lakiesha? In both instances, we distort the accuracy of the individual student's grade for any one standard.

Some collaborative projects may provide opportunities to determine individual learning regarding a specific learner outcome, but they are rare. To be accurate, then, we must assess students outside the group project to see what each one takes away from the experience. Unless we're teaching a class on group projects, group work is only the means to an end, not the actual curriculum. For grades to be accurate and useful, they must speak only to the posted curriculum.

The Problem with Averaging

We know that averaging grades falsifies grade reports (Marzano, 2000; O'Connor, 2009, 2010; Reeves, 2010; Wormeli, 2006). Henry receives an *F* on the first test but then learns the material and receives an *A* on a new assessment of the same material; unfortunately, the average of these two, a *C*, is recorded in the grade book. This is not an accurate report of Henry's newfound proficiency in the topic. If we trust the new test as a valid indicator of mastery, Henry's earlier performance is irrelevant.

Although this example uses two grading extremes (*A* and *F*), averaging grades, no matter the distance between the two or more scores, decreases accuracy. Looking at the most consistent levels of performance over time makes for a more accurate report of what students truly know, and it provides higher correlations with testing done outside the classroom (Bailey & Guskey, 2001; Marzano, 2000; Reeves, 2010).

It's unethical and inaccurate to include in a grade digressions in performance that occur during the learning process, when a grade is supposed to report students' mastery at the end of that process. It's also inaccurate to rely solely on single-sitting assessments for the most accurate report of what students know and can do. Instead, we look for evidence over time.

The Problem with Zeroes

Determining grades using the 100-point scale is ill-suited to measuring and reporting performance against specific standards. If we're calculating grades mathematically, smaller scales with clear descriptors—such as 1.0, 2.0, 3.0, and 4.0, in which all possible scores, including 0.0, have equal skewing influence on the overall score—create a more accurate report of students' mastery. Recording a zero on a 100-point scale for a student's lack of work on an assessment not only falsifies the report of what he or she knows, but also immediately generates despair: Only a mammoth pile of perfect 100s can overcome the deficit and result in a passing *D* grade. So why bother?

When considering whether to leave a score as zero or reappoint it as a 50, 59, 60, or higher (all still in the *F* range) in an effort to equalize its skewing influence, we're really deciding among variations of *F*. Do we record the lowest, most hurtful, most unrecoverable end of the *F* range—or the most hopeful, recoverable end of that range? It's a bit silly to have varying degrees of "*F*-titude," when an *F* means "no evidence of the standard yet."

The larger question really is whether we're teaching to make sure students learn the curriculum or just presenting the curriculum and documenting students' deficiencies with it. A "gotcha!" mind-set doesn't serve our mission.

Educators who consider reappointing the zero as, say, 50 may worry that students will brag to

classmates, "You worked hard, but I did nothing and still got a 50!" But students are the first to realize that they don't get something for having done nothing. Unfortunately, some teachers invoke the compensation metaphor here, claiming that they would not pay someone $50 for a job that he or she didn't do. In that context, this is correct, of course, but the analogy has nothing to do with the problem of zeroes on the 100-point scale.

If we're required to average grades, a single missing assignment—a zero—on the 100-point scale disproportionately skews the report: 100 + 100 + 100 + 0 yields an average of 75, whereas a 100 + 100 + 100 + 50 yields an average of 87.5, which is closer to the truth of overall competency if we're aggregating all assessments equally into a single, final grade. A more accurate report, however, would declare that three standards were mastered and one was not, and there would be no overall grade. Averaging muddies the grading waters, particularly with zeroes on the 100-point scale.

Doug Reeves (2004) reminds us that a zero on the 100-point scale is six levels—six increments of 10—below a failing 60 and that this equates mathematically to a -6 on the 4.0 scale. It would be absurd to record a -6 on a 4.0 scale when Ben does not submit an assignment; it's inaccurate and unfair. Ben would have to climb six levels higher just to get even with absolute failure. This practice is senseless, and it voids a school's claim to be standards-based.

For higher accuracy and effectiveness in grading, separate nonacademic elements from academic elements on the report card. Provide separate scores for each major standard or outcome within the discipline. We must end grade averaging, and if forced to use it, we must look at the evidence of students' mastery over time. Make sure that no one grade has undue skewing influence on that average.

Accurate grades provide feedback, document progress, and inform our instructional decisions. Inaccurate grades play havoc with students' lives and our professional integrity.

Consistent

Students in the classroom of teacher *x* who achieve at the same level as students in the classroom of teacher *y* should get the same grade. Schools should strive for consistency in all their classrooms, and districts should strive for consistency in all their schools.

We can achieve consistency in three ways.

Through Clarity of Purpose

Schools have used grades for a variety of purposes: communication, self-evaluation, sorting and selecting, motivation, and program evaluation (Guskey, 1996)—and therein lies the problem. Some teachers emphasize one purpose, and some emphasize another. Consequently, they use different criteria for determining grades, which can result in students who achieve at the same level receiving different grades.

To achieve consistency, schools and districts must achieve consensus about the primary purpose of grades and then publish a purpose statement that is available to all. Our premise here is that "the primary purpose of...grades [is] to communicate student achievement to students, parents, school administrators, postsecondary institutions, and employers" (Bailey & McTighe, 1996, p. 120).

Through Performance Standards

"What is good?" and "How good is good enough?" are ultimately what assessment and grading are all about, so defining the performance standards clearly, making them available to all, and ensuring that everyone understands them are essential steps to achieving consistency in grading.

A pure standards-based system would have only two levels of performance—proficient or not proficient. However, at most grade levels we may want to identify additional levels, such as above proficient, below but close to proficient, and well below proficient. This would result in a four-level system. Although there is no one right number of levels, fewer than 10 is advisable because there's a limit to how well the English language can describe different levels and how well teachers, students, and parents could understand the differences among them.

The right number of levels is a lot closer to 2 than to 100—which is why we should eliminate the percentage system because it's incompatible with a standards-based system. The two most highly regarded high school programs in the world only use levels—advanced placement uses five levels and the International Baccalaureate uses seven. Level-based systems should become the norm.

Once there's agreement on the number of levels, schools and districts need to develop and publish clear generic descriptions of each. These would then form the basis for the performance standards used in the classroom—marking schemes, rubrics, exemplars, and so forth.

Teachers must also have frequent opportunities to collaboratively assess student work so they develop common understanding of the performance standards. A common frame of reference decreases the subjective, relative, and inferential nature of grading and helps departments and grade levels recalibrate their common expectations when these expectations drift over time.

Through Clear Policies and Procedures

According to Carifio and Carey (2009), "Many schools lack a coherent and uniform grading policy, resulting in extensive variations in student assessment from teacher to teacher, and even between students taking the same course with the same teacher" (p. 25). It's therefore crucial that all schools and districts have public, published policies and procedures that all teachers are expected to follow and for which they can be held accountable if students, parents, or administrators identify concerns with their grading practices.

Meaningful

Let's look at three hypothetical report cards. John's report card indicates he got a *B* in mathematics. Brian's report card indicates he got a *B* in number sense, a *C* in calculation, and an *A* in measurement. Marilyn's school uses a four-level scale: 4 for *excels*, 3 for *proficient*, 2 for *approaching proficiency*, and 1 for *well below proficiency*. Her report card indicates the following[1] :

Number Sense

Identifies place value to 1000s:	4
Reads and writes common fractions:	3
Reads whole numbers through four digits:	3

Writes whole numbers through four digits: 3

Orders and compares whole numbers through four digits: 1

Computation

Addition: 4

Subtraction: 3

Multiplication: 3

Division: 1

Uses calculator to add or subtract numbers with 4 or more digits: 2

Estimation skills: 4

It's obvious that Marilyn's report card has much more meaningful information than John's and Brian's report cards do and that Brian's report card provides more meaningful information than John's does. Single-subject grades—John's *B* in math—provide little useful information. Providing standards-based grades makes grades meaningful because they clearly show the student's areas of strength and areas that need improvement. This type of standards-based grading should be the norm from kindergarten to grade 12 (and beyond!) because it gives students, parents, and teachers the valuable information they need to help students achieve at higher levels.

Teachers traditionally have organized their grade books with categories for tests, projects, and assignments; the base has been assessment methods or activities. However, in standards-based systems, the base should be some structure coming from the standards. The level of specificity may vary from grade level to grade level and from subject to subject. The categories may be broad, as illustrated by Brian's report card, or specific, as illustrated by Marilyn's report card.

Supportive of Learning

Grades are small symbols used as shorthand for much larger descriptors. Contrary to the emotional baggage so often applied to each one, they are not full descriptors themselves. To support students' learning, they must be informative. We're mindful of each symbol's purpose in the learning process and, in particular, whether they refer to formative or summative assessments.

Because we don't want to diminish the powerful effect that formative assessments bring to students' learning, we use scores only from summative assessments to determine grades. Formative assessment uses symbols or narrative commentaries that are *not* included in determining grades.

Effective assessment is revelatory; it reveals the student's story. Students need a safe place to tell that story and receive helpful feedback on its unfolding. For that feedback to be useful, we limit judgment and evaluation. We reflect back to students how they performed on assessments and then help them compare their performances to standards of excellence set for those tasks. If we grade the formative steps that students take as they wrestle with new learning, every formative assessment becomes a final judgment, with no chance for revision and improvement. Feedback is diminished, and learning wanes.

To be useful then, formative and summative reports must be distinct from one another. We set up grade books in two sections, *formative* and *summative*; or we label each assessment with an "F" or "S"; or we color-code assessments accordingly, such as red for *formative* and green for *summative*. An assessment is formative or summative depending on when we give it and how we use the resulting data.

Most formative assessments provide descriptive feedback to students, followed by opportunities to revise in light of that feedback and be assessed and accredited anew. We want to protect that learning cycle as much as we can; most professionals follow this kind of development cycle throughout their careers.

Summative assessments, on the other hand, are for evaluative declarations and sorting students. They do not offer much in the way of feedback and opportunities for revision and reassessment. The use of formal letter grades and judgment symbols are appropriate for such assessments.

Interestingly, if we're living up to the promise of teaching every student, not just the easy ones, we could turn all summative assessments into formative ones. The only reason students can't redo a final exam, project, or standardized test after they receive feedback and revise their learning is that someone in a policy-making capacity declared it so—not because it's bad pedagogy.

We Owe Them This

When did we drift into grades of unquestioned provenance becoming the legitimate currency for the next generation? And why do we succumb to the notion that because something is easy to calculate it must be pedagogically sound?

With accountability measures on the rise and both businesses and colleges questioning the validity of the modern high school diploma, grading and standards are now under intense scrutiny. We can no longer afford the mind-set "You do your thing, and I'll do my thing" when it comes to either. We need honest, useful reports of student performance on standards and outcomes. Our students' futures depend on it.

References

Bailey, J. M., & Guskey, T. R. (2001). *Developing grading and reporting systems for student learning.* Thousand Oaks, CA: Corwin.

Bailey, J. M., & McTighe, J. (1996). Reporting achievement at the secondary school level: What and how? In T. R. Guskey (Ed.), *Communicating student learning: ASCD yearbook 1996.* Alexandria, VA: ASCD.

Carifio, J., & Carey, T. (2009). A critical examination of current minimum grading policy recommendations. *The High School Journal, 93*(1), 23–37.

Guskey, T. (1996). Reporting student learning: Lessons from the past—prescriptions for the future. In T. R. Guskey (Ed.), *Communicating student learning: ASCD yearbook 1996.* Alexandria, VA: ASCD.

Marzano, R. (2000). *Transforming classroom grading*. Alexandria, VA: ASCD.

O'Connor, K. (2009). *How to grade for learning: Linking grades to standards*. Thousand Oaks, CA: Corwin.

O'Connor, K. (2010). *A repair kit for grading: 15 fixes for broken grades*. Boston: Pearson.

Reeves, D. B. (2004). The case against the zero. *Phi Delta Kappan, 86*(4), 324–325.

Reeves, D. B. (2010). *Elements of grading*. Bloomington, IN: Solution Tree.

Wormeli, R. (2006). *Fair isn't always equal: Assessment and grading in the differentiated classroom*. Portland, ME: Stenhouse.

Endnote

1 The categories used here come from Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning* (p. 289). Boston: Pearson.

Ken O'Connor is an independent consultant and author who specializes in issues related to communicating student achievement; kenoc@aol.com. **Rick Wormeli** is a national education consultant, veteran educator, and author; rwormeli@cox.net.

KEYWORDS

Click on keywords to see similar products:
student assessment

Requesting Permission

- For **photocopy, electronic and online access**, and **republication requests**, go to the Copyright Clearance Center. Enter the periodical title within the **"Get Permission"** search field.
- To **translate** this article, contact permissions@ascd.org