



# Chapter 1

## Setting the Stage

*School improvement expert Bob Marzano asks, “Why [w]ould anyone want to change current grading practices? The answer is quite simple: grades are so imprecise that they are almost meaningless.”*

—Marzano, 2000, p. 1

Every state in the United States, every province in Canada, and almost every jurisdiction in most other countries now has educational content *standards*—public, published statements of the expected outcomes of learning; that is, what students are expected to know, understand, and be able to do. The primary goal of a standards-based system is for all students to “meet standards”; that is, to be competent or proficient in every aspect of the curriculum. The key to reaching this goal is to evaluate every student’s achievement using similar criteria, consistently applied at all levels.

The two essential questions that all educators should ask about their grades are, “How confident am I that the grades students get in my classroom/school/district are accurate, meaningful, and consistent, and that they support learning?” and “How confident am I that the grades I assign students accurately reflect my school’s/district’s published content standards and desired learning outcomes?” In most schools/districts the answers to these questions, especially at the middle and high school levels, range from “not very” to “not at all.” Because of this I believe that, very often, grades are “broken” and that teachers and schools/districts need a “repair kit.” I offer such a kit here in the form of 15 Fixes that teachers and administrators alike can apply to repair broken grading systems.

Clear evidence of this broken condition can be found in the following example, from a blog on the *Washington Post* website on July 27, 2008. The previous day Jay Mathews had written an article about Matt, a student who had been asked to leave Thomas Jefferson High School of Science and Technology in Fairfax County, Virginia, because he had not attained the 3.0 GPA school policy requires that students maintain at the end of their sophomore year.

Before I start, I would like readers to know that I am currently a rising junior at TJ (Thomas Jefferson High School for Science and Technology in Fairfax County, VA) so my views will have some bias in it (although I will do my best to minimize it). . . .

For example, I recently finished an AP chemistry course. This class was coated with extra credit, and with a 0.5 boost on GPA's, my teacher was basically handing out either a 4.0 or 4.5 for chemistry. Another AP chemistry teacher is very strict about extra credit, and rarely gives any. His course is equally challenging to my course. I got a B with extra credit, without it, I would have a C or D.

I consider myself a below par TJ student. Every class I have taken, I've always been below the class average (I even took Mrs. G. . . 's anthropology class, anyone you ask will tell you that many people take that class for the easy A. I got a B+). My current GPA is around 3.4 and *I have to admit, at least half of my sophomore grades are inflated by at least a whole letter.* If I had Matt's (the TJ student with a 2.8 GPA) combination of teachers, I would have been likely to be in his place. *The large amount of chance factors heavily on ones [sic] GPA.* (emphasis added) (zhengfranklin, 2008)

This high-stakes grading environment appears to lack consistency and accuracy, as teachers teaching the same course obviously have very different grading procedures. As Jay Mathews points out, this results in grades being *a matter of chance* as to which teacher the computer assigns a student—and this school is frequently ranked as the number one high school in the United States.

As the first “essential question” indicates, effective grades need to meet four overarching criteria for, or keys to, success: they must be accurate, meaningful, and consistent, and must support learning. I define each of these keys here and then weave them into each of the 15 Fixes throughout the rest of this book. I believe that most teachers, students, and parents would agree that these are reasonable and necessary expectations; disagreements over how to achieve them within the grade itself are at the root of the debate about grading.

First and foremost, grades need to be *accurate* reflections of student achievement. Inaccurate grades lead to poor decisions being made by and about any student whose grades are used as the basis

of those decisions. When determining grades, many teachers continue the traditional practice of combining a large amount of evidence/data into a single summary symbol. This may involve literally hundreds of decisions; if even one is wrong the grade inaccurately reflects student achievement. Inaccurate grades most commonly result from teachers determining them by blending achievement with behaviors (effort, participation, adherence to class rules, etc.) (Fix 1), poor-quality assessment (Fix 10), and inappropriate use of the mean (average) in combining data (Fix 11). For grades to be “fixed,” each of these practices (and others, discussed in Fixes 4, 5, 8, 9, and 12) needs to be eliminated.

Grades need to be *meaningful*. They must communicate useful information to students and to everyone interested in or needing to know about their learning. Traditionally, teachers have collected evidence using various assessment methods and have organized their gradebooks by type of evidence such as tests, projects, and assignments. So, the grading link to learning outcomes has been tentative at best. The “fix” needed for grades to be meaningful is that they must directly reflect specified learning goals. This requires that teachers set up and organize their gradebooks around those goals or standards—not simply summarize multiple marks into a single grade, or organize grades by the date administered, type of assignment or activity, or type of test—by using the standards or some organizational structure arising from or related to the standards (Fix 7). The evidence categories for mathematics, for example, may include “develops and uses number strategies,” “compares and orders whole numbers to 100,” and “uses estimation strategies.” The evidence structure for English may use strands such as reading, writing, listening, speaking, language, and literature.

Grades need to be *consistent* across teachers. The grades students receive should not be a function of whether they are in teacher X’s or teacher Y’s class. The question, “How good is good enough?” needs to be the same from classroom to classroom; that is, performance standards need to be the same from teacher to

teacher. Students achieving at the same level should get the same grade regardless of context. This clearly is not the case in schools where some teachers are identified as “hard” and others labeled as an “easy A,” as in the example from Thomas Jefferson High School. This should *never* be acceptable. To “fix” grades, especially in standards-based systems, it is at minimum essential that all teachers in every school teaching the same grade or same subject/course should determine grades in similar ways and apply similar or the same performance standards. This consistency in the meaning of grades should be systemic at all levels—school, district, and, ideally, state/province.

Grades need to *support learning*. Students and parents need to understand that achieving in school is not only about “doing the work” or accumulating points. When teachers assign a point value to simply turning in work, or put a mark or number on everything students do and use every number when calculating the grade, the message sent to students is clear: success lies in the quantity of points earned. Any intended message about valuing the *quality* of the learning is blurred. We want students to understand that school is about learning. Grades are artifacts of learning; as such, they should reflect student achievement only (Fixes 2, 3, and 6).

Grades also support learning when the purpose of each assessment is clear. *Formative* assessments are designed to help students improve, and in almost all cases should not be used to determine grades. *Summative* assessments are designed to measure student achievement, and “are used to make statements of student learning status at a point in time to those outside the classroom” (Stiggins et al., 2004, p. 31). With some limited exceptions, only evidence from summative assessments should be used when determining grades (Fix 13). We also must allow new evidence to replace old evidence when it is clear that a student knows or can do something today that they didn’t or couldn’t previously (Fix 14). Finally, and perhaps most importantly, for grades to support learning, we must learn how to involve students in the grading process (Fix 15).

## Key Definitions

One problem is that the terms *marks* and *grades* are often mistakenly used as synonyms, although each involves very different processes. A teacher looking at a single assessment and deciding whether a student should get 7 or 8 out of 10, or a 3 or 4 on a rubric, is doing something very different than when that teacher is looking at the evidence accumulated over a grading period and deciding whether that student gets an A or a B (or whatever summary symbols are used) on their report card. To avoid confusion, we use the following definitions throughout this book (note, however, that the sources quoted herein may not necessarily follow these definitions):

- A *mark* or *score* is the number (or letter) given to any student test or performance that may contribute to the later determination of a grade.
- A *grade* is the symbol (number or letter) reported at the end of a period of time as a summary statement of student performance.

## Purpose(s) for Grades

Traditionally, grades have served a number of purposes—communication, fostering student self-assessment, sorting and selecting, motivation and punishment, and teaching/program evaluation (Guskey, 1996a). As Brookhart (2004, p. 21) points out, “It is very difficult for one measure to serve different purposes equally well.” She also states, “The main difficulty driving grading issues both historically and currently is that grades are pressed to serve a variety of conflicting purposes” (p. 31). For example, for communicating effectively in a standards-driven environment where many students are succeeding, we need to be communicating the highest possible achievement in the narrowest possible range—all students are successful. However, for sorting and selecting these same students we need to spread them along the widest possible range, thus ranking some high and some low. These two purposes clearly can be in

conflict. Bailey and McTighe (1996, p. 120) state that “the *primary purpose . . .* of grades [is] to *communicate student achievement* to students, parents, school administrators, post-secondary institutions and employers.” Brookhart (2004, p. 5) suggests, “*Secondary purposes* for grading include providing teachers with information for instructional planning, . . . and providing teachers, administrators, parents, and students with information for selection and placement of students” (emphasis added).

A central premise of this book is that, at the district and school levels, there must be a shared vision of the primary purpose of grades. I believe that primary purpose to be communication about achievement, with *achievement* being defined as performance measured against accepted published standards and learning outcomes.

## Underpinning Issues

There are three underpinning issues we must consider before addressing the specifics of how to determine grades. They are fairness, motivation, and objectivity and professional judgment.

### Fairness

In education we have tended to think of *fairness* as uniformity. All students have been required to do the same assessments in the same amount of time and their grades have been calculated in the same way from the same number of assessments. But students are different in many different ways, and so treating them the same can actually be unfair. Patterson (2003, p. 572) points out that “fair does not mean equal; yet, when it comes to grading, we insist that it does.” Fairness is much more about equity of opportunity than it is about uniformity. For example, some students need to wear glasses and for equity of opportunity they wear their glasses when they need them; for fairness we do not say, “You are doing a test today but you cannot wear your glasses because everyone is not wearing glasses” or “Some students in this class need glasses, so you will all wear them (whether you need them or not).”

This concept has been captured in the following statement from the policy about provincial testing in Manitoba. All teachers and jurisdictions would be serving their students well if they had a similar statement in their school/classroom assessment/grading policy:

All students are given an equal opportunity to demonstrate what they know and can do as part of the assessment process. Adaptations . . . are *available* for students including students with learning or physical *disabilities*, to allow them to demonstrate their knowledge and skills, provided that the adaptations do not jeopardize the integrity or content of the test. (Manitoba Education, Citizenship and Youth, 2006, p. 1, emphasis added)

The italics emphasize that, for fairness, “adaptations” should not be limited to students who have been specifically identified as needing, for example, more time to complete a test/exam.

## Motivation

Grades are often *extrinsic* motivators, meaning that their power to influence student behavior derives from outside the student. Many teachers—and parents, grandparents, and other adults—have used grades as extrinsic motivators (“Everyone who gets an A on this quiz can skip the next homework assignment”; “Get a B or better on that test or you can’t go to the concert”). However, this use of grades is not always effective or appropriate. Grades certainly motivate successful students, at least some of the time. But they are definitely not motivators for all students, such as those who get grades that are lower than they expect or think they deserve. For these students, grades in fact often act as *de*-motivators. Many schools and school districts have mission or belief statements that state their desire to develop students who are “independent, self-directed, lifelong learners.” To achieve this goal students need to be *intrinsically* motivated, meaning that their desire to achieve and improve must arise from within themselves. Intrinsic motivation is clearly in conflict with the use of grades as extrinsic motivators. Thus, as we think about our

current and future grading practices, it is important that we examine and apply our knowledge and beliefs about what does and does not motivate students.

Consider this quote from Nora Rowley, the fifth-grade student who is the main character in Andrew Clements's *The Report Card*:

Most kids never talk about it, but a lot of the time bad grades make them feel dumb, and almost all the time it's not true. And good grades make other kids think they're better, and that's not true either. And then all the kids start competing and comparing. The smart kids feel smarter and get all stuck-up, and the regular kids feel stupid and like there's no way to catch up. And the people who are supposed to help kids, the parents and the teachers, they don't. (Clements, 2004, pp. 72–73)

Clements, through Nora, makes it clear that he believes that not only do grades not motivate many students, but also that they can actually damage both student attitudes toward learning and relationships among students. Both in and out of school we provide elaborate systems of rewards and punishments in the belief that this will lead to more of those behaviors deemed desirable and less deemed undesirable. But the research on motivation shows that continued use of extrinsic motivators leads to two main results. First, extrinsic motivators increase students' focus on the reward or punishment rather than on the desired behavior. Second, they give rise to the need to continuously increase the amount of the reward or punishment to elicit the desired behavior (Covington & Manheim Teel, 1996; Gathercoal, 2004; Ginsberg, 2004; Kohn, 1993; Marshall, 2001a; Rogers, Ludington, & Graham, 1998; Szatanski & Taafe, 1999). Thus it is inappropriate to use grades as extrinsic motivators, either to reward desired behavior or to punish undesired behavior. The primary "reward" for learning should be intrinsic—the positive feelings that result from success. As Stiggins notes, "those who experience . . . success gain the confidence needed to risk trying. . . . Students who experience . . . failure, lose confidence in themselves,

stop trying, and [fail] even more frequently. . . . As it turns out, *confidence is the key to student success in all learning situations*” (2001, p. 43, emphasis in original). Actual success at learning, then, is the single most important factor in (intrinsic) motivation, and it is important to recognize that success is relative—success for each individual is seeing oneself getting better.

Additionally, teachers have other tools available to help them change student behavior. As Marshall (2001b, p. 9) points out, “the most effective ways to change behaviors are: (1) using noncoercion, (2) prompting the person to self-assess, and (3) if authority is necessary, having the student own the consequence. When a consequence is imposed the student feels the victim. When the consequence is elicited, the student owns it and grows from the decision.”

The best classroom practices maximize intrinsic motivation and minimize extrinsic motivation. Teachers in these classrooms help students to the critical understanding that “30 years from now, it won’t matter what grades you got. What will matter is what you learned and how you used it.”\*

Daniel Pink (2009) writes in his recent book *Drive* that he believes that this is now also true in the world of work; the connections between motivation in the work world and in schools is examined in an interview with Pink in the May/June 2010 issue of *Edge*. He states,

The reason to do this (emphasize intrinsic motivation) is not to be soft-hearted about it or be nicer about it. The reason to do it is because science says that if you really want high performance, particularly on those 21st century tasks, those old 19th century motivators aren’t going to get you there. We need to fundamentally rethink things, not to be nicer and kinder, but to be more effective and productive. I think this approach has the ancillary benefit of being more humane, but I think it also has a very hard-headed benefit of being much more effective, much more productive, and will actually lead

\*From a poster seen on the wall of a high school cafeteria in Council Bluffs, Iowa. Source otherwise unknown.

us to the sort of accountability that many people, including myself, think is essential. . . . The big thing is getting beyond the folklore of what really motivates people into the science of what motivates people. (Phi Delta Kappa International, 2010, pp. 5–6)

## Objectivity and Professional Judgment

Teachers often say that they are striving to be as objective as possible in their assessment and grading. In my experience, they most often mean that they are trying to be *consistent* in evaluating student work. Such a process in fact involves subjective judgment. The only aspects of learning that can be assessed objectively are such elements as the correctness of factual content, spelling, and calculation.

Assessments themselves are designed subjectively. Teachers create assessments based on their professional judgment of what is to be assessed and how—a subjective process. We need to acknowledge this and not apologize. As Wiggins (2000, n.p.) notes, “All scoring by human judges, including assigning points and taking them off math homework is subjective. The question is not whether it is subjective, but whether it is defensible and credible. The Advanced Placement and International Baccalaureate assessments are subjective and yet credible and defensible, for example. So-called objective scoring is still subjective test writing.” Thus the real issues are accuracy and consistency, more than objectivity versus subjectivity. We need to develop approaches to help teachers both assess and grade more accurately and consistently. One key to accomplishing this is shared understanding of performance standards—agreement on “how good is good enough?” Another is unified approaches to determining grades at the school or district level.

The problem as identified by an assistant superintendent in a Wisconsin school district is that in grading “every teacher sees himself or herself as an independent contractor and they shouldn’t be” (personal communication, n.d.). What is needed is a set of guidelines such as the 15 Fixes provided in this book. Making these Fixes part of district or school policy and providing teachers frequent

opportunities both for professional learning and dialogue about these guidelines and to carry out shared marking to arrive at a common understanding of performance standards will greatly enhance the probability of consistent grading across teachers and classrooms. Grading must not be a private practice; it must be a shared practice.

## **Student Involvement**

Over the last few years it has become increasingly clear that student involvement in teaching/learning and in assessment and communication can make significant contributions to improved achievement and positive attitudes about learning/school. This issue is so significant I incorporate suggestions about it into many of the Fixes. Fix 15 is a synthesis and summary of these ideas as presented throughout the rest of the book.

## **The 15 Fixes**

The 15 Fixes appear in Figure 1.1. They are organized into four categories—fixes for distorted achievement, fixes for low-quality or poorly organized evidence, fixes for inappropriate grade calculation, and fixes to support learning. We discuss each Fix in turn in the following chapters.

**Figure 1.1** The 15 Fixes

---

**Fixes for Practices That Distort Achievement**

1. Don't include student behaviors (effort, participation, adherence to class rules, etc.) in grades; include only achievement.
2. Don't reduce marks on "work" submitted late; provide support for the learner.
3. Don't give points for extra credit or use bonus points; seek only evidence that more work has resulted in a higher level of achievement.
4. Don't punish academic dishonesty with reduced grades; apply other consequences and reassess to determine actual level of achievement.
5. Don't consider attendance in grade determination; report absences separately.
6. Don't include group scores in grades; use only individual achievement evidence.

**Fixes for Low-Quality or Poorly Organized Evidence**

7. Don't organize information in grading records by assessment methods or simply summarize into a single grade; organize and report evidence by standards/learning goals.
8. Don't assign grades using inappropriate or unclear performance standards; provide clear descriptions of achievement expectations.
9. Don't assign grades based on student's achievement compared to other students; compare each student's performance to preset standards.
10. Don't rely on evidence gathered using assessments that fail to meet standards of quality; rely only on quality assessments.

**Fixes for Inappropriate Grade Calculation**

11. Don't rely only on the mean; consider other measures of central tendency and use professional judgment.
12. Don't include zeros in grade determination when evidence is missing or as punishment; use alternatives, such as reassessing to determine real achievement, or use "I" for Incomplete or Insufficient Evidence.

**Fixes to Support Learning**

13. Don't use information from formative assessments and practice to determine grades; use only summative evidence.
14. Don't summarize evidence accumulated over time when learning is developmental and will grow with time and repeated opportunities; in those instances, emphasize more recent achievement.
15. Don't leave students out of the grading process. Involve students; they can—and should—play key roles in assessment and grading that promote achievement.

