DOUGLAS REEVES

ELEMENTS *of* GRADING

SECOND EDITION

A Guide to Effective Practice

Solution Tree | Press

Chapter 3

THE IMPACT OF FEEDBACK ON ACHIEVEMENT

Although grading policies can be the subject of deeply held opinions, debates about grading are more constructive if we first agree on two important premises. First, we should be willing to agree that grading is a form of feedback. Second, we should be willing to agree that feedback is a very powerful instructional technique—some would say the most powerful—when it comes to influencing student achievement.

Evaluating the Evidence

Let's look at the evidence. John Hattie's (2009) synthesis of more than eight hundred meta-analyses evaluates the relative impact of many factors, including family structure, curriculum, teaching practices, and feedback on student achievement. The measurement that Hattie uses is effect size, or, simply put, the effectiveness of particular interventions. The impact of an effect size of 0.4 is, according to Hattie, about one year of learning. Therefore, any instructional or leadership initiative must at least pass this threshold. Many factors are statistically significant, as the following list will show. But statistical significance and practical significance are two different elements. Because of the overwhelming burdens on the time and resources of every school (Reeves, 2011a), it makes little sense to invest in

initiatives that fail to cross the 0.4 level in effect size. An effect size of 1.0, Hattie suggests, would be blatantly obvious, such as the difference between two people who are 5 feet 3 inches (160 cm) and 6 feet (183 cm) in height—a difference clearly observable.

Even small effect sizes can be meaningful, particularly if they are devoted to initiatives that save lives. For example, Robert Rosenthal and M. Robin DiMatteo (2001) demonstrate that the effect size of taking a low dose of aspirin in preventing a heart attack is 0.07—a small fraction of a standard deviation—yet this translates into the result that thirty-four out of every one thousand people would be saved from a heart attack by using a low dose of aspirin on a regular basis.

The use of the common statistic for effect size helps busy teachers and school administrators evaluate alternative strategies and their impact on achievement compared to variables outside teachers' and students' control. For example, some of Hattie's findings include the influence of the following on student achievement (Hattie, 2009).

- Preterm birth weight (0.54)
- Illness (0.23)
- Diet (0.12)
- Drug use (0.33)
- Exercise (0.28)
- Socioeconomic status (0.57)
- Family structure (0.17)
- Home environment (0.57)
- Parental involvement (0.51)

Most teachers would view these factors as outside of their control, although some would certainly argue that schools can do a better job of influencing diet, drug use, exercise, and parental involvement. During the eighteen hours every day that students are not in school, students and families make many decisions that influence learning in significant ways. But how important are these decisions compared to the variables that teachers and school administrators can control?

The Importance of Feedback

The effectiveness of any recommendation regarding teaching and education leadership depends on the extent to which the professional practices of educators and school leaders have a greater impact on students than factors that are beyond their control. The essential question is, Will this idea have a sufficient impact in helping students overcome any negative influences they face outside of school?

Fortunately, Hattie (2009) answers that question with a resounding affirmative response. He finds a number of teaching and leadership practices that, measured in the synthesis of meta-analyses, are more powerful than personality, home, and demographic factors when considering their impact on student achievement. Examples include teacher-student relationships (0.72), professional development (0.62), teacher clarity (0.75), vocabulary programs (0.67), creativity programs (0.65), and feedback (0.73).

Certainly, Hattie is not the first scholar to recognize the importance of feedback on student achievement. His findings are completely consistent with Robert Marzano's (2007, 2010) conclusions that accurate, specific, and timely feedback is linked to student learning. Thanks to Hattie's research, however, we can now be more precise than ever about how important it is. We can say that, based on the preponderance of evidence from multiple studies in many cultural settings, feedback is not only more important than most other instructional interventions but is also more important than socioeconomic status, drug use, nutrition, exercise, anxiety, family structure, and a host of other factors that many people claim are overwhelming. Indeed, when it comes to evaluating the relative impact of what teachers and education leaders do, the combined use of formative evaluation and feedback is the most powerful combination that we have. If we understand that a grade is not just an evaluation process but also one of the most important forms of feedback that students can receive, Hattie's conclusion should elevate the improvement of grading policies to a top priority in every school.

Hattie (2009) also encourages a broadly based view of feedback, including feedback not only from teachers to students but also from teachers to their colleagues. We should recall that, as a fundamental

ethical principle, no student in a school should be more accountable than the adults, and thus our feedback systems must be as appropriate for teachers and leaders as they are for students. Similarly, our standards for administrators, board members, and policymakers must be at least as rigorous as those we create for fourth graders. If that statement seems astonishing, then I invite you to obtain a copy of the fourth-grade academic standards for your area and lay beside them the standards that are officially endorsed for policymakers, such as legislators, members of parliament, members of Congress, or other educational authorities. You can then decide which standards are more demanding.

The Evidence–Decision Gap

It is therefore mystifying that a strategy with so great an impact on student achievement as feedback remains so controversial and inconsistent. It is as if there was evidence that a common consumer practice created an environmental disaster, but people ignored it and persisted in the destructive practice. Of course, that is hardly a hypothetical example, as our national habits—such as persistent use of bottled water, dependence on gas-guzzling cars, and appetite for junk food—illustrate. Rather than embrace the evidence and use filtered tap water, take public transportation, and eat fresh vegetables, we often choose the convenient alternatives that are less healthy for our families and the planet.

In sum, our greatest challenge is how to transform what we know into action. Indifference to research, though also present in medicine, business, and many other fields (Pfeffer & Sutton, 2006a), is particularly striking in education. An alarming example is the persistent use of retention and corporal punishment. In both cases, decades of evidence suggest that these "treatments" are inversely related to student learning. Retention does not encourage work ethic and student responsibility but only creates older, frustrated, and less successful students (Hattie, 2009). Corporal punishment does not improve behavior but legitimizes violence and increases bullying and student misbehavior (Committee on School Health, 2000). Nevertheless, politicians from all parties have excoriated social promotion and urged retention in a display of belligerent indifference to the evidence. More disturbingly,

nineteen states and many other nations continue to permit corporal punishment decades after the evidence concluded it was counterproductive (Strauss, 2014).

Equipped with rich literature on the theory and practice of change, educators and school leaders should be fully capable of acknowledging error, evaluating alternatives, testing alternative hypotheses, and drawing conclusions that lead to better results. Instead, personal convictions that are not only antiquated but maybe even dangerous guide decision-making processes. We can be indignant about the physicians of the 19th century who were unwilling to wash their hands, but when the subject turns to education policies, we sometimes elevate prejudice over evidence.

Before we consider what quality feedback is, let us be clear about what feedback is not. Feedback is not testing.

Distinguishing Feedback From Testing

Consider two classrooms, both burdened by large class sizes and students with a wide range of background knowledge and skill levels. The role of the teacher in the first class is to deliver what, as a matter of school-system policy, has been described as a "guaranteed curriculum." Administrators know that the curriculum is delivered because teachers list the instructional objectives on the board and post the details of the lesson plan supporting those objectives next to the door, where visiting leaders can easily inspect them. In this class, the most important feedback that students and teachers receive is on the annual test administered every spring. This feedback is very detailed, as it determines the success and failure of not only individual students but also the entire school, perhaps the entire school system. Moreover, external companies have established elaborate statistical formulas that give feedback to individual teachers, measuring the degree to which each teacher is adding value to each student.

When comparing students over three years, these analyses conclude that teachers whose students show gains in test scores have added value to their students, whereas teachers whose students do not make such gains have failed to add value. So ingrained is this sort of analysis that in the United States, one of the conditions states

must meet in order to be competitive for federal funds is the commitment to link teacher evaluation to annual measures in student performance.

There is no question that annual tests are important, if by important we mean that decisions involving the lives of students, teachers, and school administrators, along with billions of taxpayer dollars, are influenced by those tests. Ask the teacher and students in the first class how they know when they are succeeding, and the answer is, almost uniformly, "We'll know when we get our state test results back." However, the question at hand is whether these test results really provide feedback.

The second class is no less rigorous than the first. Indeed, it can be argued that this class is more rigorous. The teacher provides informal feedback to students every day, and each week students update their learning logs to identify where they are with respect to their learning targets and next steps for moving forward. Students, along with the teacher, are continuously assessing their learning but not with a single standardized test. Moreover, the teacher in the second class assesses skills that are never tested by the state, including collaboration, critical thinking, creativity, and communication. This teacher is not assessing less but assessing more to prepare students not only for the state test but also for the broader requirements students will encounter in the years ahead.

Reconsidering Feedback

In her landmark work comparing high- and low-performing nations and high- and low-performing state education systems, Linda Darling-Hammond (2010) comes to an astonishing and counterintuitive conclusion. Since the 1980s, the three exemplars she considers—Singapore, South Korea, and Finland—made significant progress according to international education comparisons over the next three decades. More than 90 percent of the students in these countries graduate from high school, and large majorities go to college—"far more than in the much wealthier United States" (p. 192), Darling-Hammond concludes. Detailed field observations reveal the rich, nuanced feedback that students and teachers receive daily and can apply immediately.

"Wait," you may say. "Don't Asian countries like South Korea and Singapore also have a test-focused environment? Aren't those the examples that we tried to emulate to improve our academic performance in mathematics and science?" In fact, this does not comport with Darling-Hammond's (2010) evidence. These successful nations:

> *eliminated examination systems* that had previously tracked students for middle schools and restricted access to high school. Finland and Korea now have no external examinations before the voluntary matriculation exams for college. In addition to the "O" level matriculation examinations, students in Singapore take examinations at the end of primary school (grade 6), which are used to calculate value-added contributions to their learning that are part of the information system about secondary schools. These examinations require extensive written responses and problem solving, and include curriculum-embedded projects and papers that are graded by teachers. (p. 192, emphasis in original)

Effective education systems certainly use some system-level examinations, but notice the important distinctions. In these examples, even national examinations include deep teacher involvement and, therefore, offer the opportunity for feedback that is far more nuanced than a simple score. Most importantly, the vast majority of feedback is in the daily interactions between students and teachers, not from test scores administered at multiyear intervals. Perhaps the most important consideration is how teachers and students evaluate their own success. While annual high-stakes testing leaves students and teachers wondering about their success ("We'll know how we're doing when we see the scores at the end of the year"), a system characterized by effective feedback offers a dramatically different view.

Darling-Hammond (2010) observes the dramatic difference between the feedback as testing model and the feedback as breathing model, with the latter characterized by feedback integral to the minute-to-minute reality of the classroom. The following words are not from a veteran teacher, nor are they from the graduate of a top-tier teacher-preparation program with several years of intensive mentoring. They are the words of a prospective teacher who was

fortunate enough to see Darling-Hammond's (2010) fieldwork but had not yet spent a day in the classroom. This teacher says:

> For me the most valuable thing was the sequencing of the lessons, teaching the lesson, and evaluating what the kids were getting, what the kids weren't getting, and having that be reflected in my next lesson . . . the "teach-assess-teach-assess-teach-assess" process. (as cited in Darling-Hammond, 2010, p. 223)

Bridget Hamre of the University of Virginia Curry School of Education notes that "high-quality feedback is where there is a back-and-forth exchange to get a deeper understanding" (as cited in Gladwell, 2009, p. 326). Bob Pianta, dean of the Curry School, reports on what a team he led observed in a class with high levels of interactive feedback:

> "So let's see," [the teacher] began, standing up at the blackboard. "Special right triangles. We're going to do practice with this, just throwing out ideas." He drew two triangles. "Label the length of the side, if you can. If you can't, we'll all do it." He was talking and moving quickly, which Pianta said might be interpreted as a bad thing, because this was trigonometry. It wasn't easy material. But his energy seemed to infect the class. And all the time he offered the promise of help. If you can't, we'll all do it.
>
> In a corner of the room was a student named Ben, who'd evidently missed a few classes. "See what you can remember, Ben," the teacher said. Ben was lost. The teacher quickly went to his side: "I'm going to give you a way to get to it." He made a quick suggestion. "How about that?" Ben went back to work. The teacher slipped over to the student next to Ben and glanced at her work. "That's all right!" He went to a third student, then a fourth. Two and a half minutes into the lesson—the length of time it took [a] subpar teacher to turn on the computer—he had already laid out the problem, checked in with nearly every student in the class, and was back at the blackboard to take the lesson a step further.

> "In a group like this, the standard MO would be: he's at the board, broadcasting to the kids, and has no idea who knows what he's doing and who doesn't know," Pianta said. "But he's giving individualized feedback. He's off the charts on feedback." Pianta and his team watched in awe. (as cited in Gladwell, 2009, p. 329)

The danger in observing an exemplary teacher is that we can relegate these experiences to the realm of mystery. Why is he such a great teacher? Some people might conclude that it must be a combination of talent, intuition, mystical insight, and a knack—he just "has it" (*it* being those amazing qualities that all exceptional teachers share). However, we might not say that about a great physician, scientist, attorney, race car driver, violinist, or basketball star. Indeed, the overwhelming evidence is that talent is not a mystery but something developed with deliberate practice (Colvin, 2008; Ericsson, Charness, Hoffman, & Feltovich, 2006). Can we apply that generalization to teaching? Here, too, the evidence demonstrates convincingly that feedback, along with other effective teaching techniques, is a skill that can be observed, applied, practiced, and improved (Lemov, 2010).

The Four Elements of Effective Feedback

As we have seen, the clear preponderance of evidence is not only that feedback is important in influencing student achievement but also is relatively more important than almost any other student-based, school-based, or teacher-based variable. It should be noted that evidence on the power of feedback is hardly restricted to the world of education. Dianne Stober and Anthony Grant (2006) and Alan Deutschman (2007) provide evidence from a wide range of environments that depend on feedback, including health care, prisoner rehabilitation, recovery from addiction, and education. Kerry Patterson, Joseph Grenny, David Maxfield, Ron McMillan, and Al Switzler (2008) add to the body of evidence, using cross-cultural examples in which people are engaged in significant and profound change, even though they cannot read or write.

In brief, it is not the provision of a data-driven, decision-making seminar that helps individuals, organizations, or communities change.

Instead, it is the ability to use feedback in clear and consistent ways. However, even the most clear and vivid feedback is useless if not applied with the FAST elements. Each of these is a necessary but insufficient condition for improvement. If information is accurate but not timely, it is unlikely to lead to any improvements. An autopsy, for example, is a marvelously accurate piece of diagnostic work, but it never restores the patient to health.

Almost every teacher I know labors to be fair, excluding any bias regarding gender or ethnicity, in their evaluations of student work, but the pursuit of fairness can impair accuracy. This is particularly true when teachers conflate a student's attitude and behavior with the quality of his or her work. Many computer programs can provide rapid feedback, but if that feedback only informs students whether their performance is correct or incorrect, they will gain little information about how to improve the thinking process that led to an incorrect response or how to sustain the analyses that led to a correct one. Specificity is a component of effective feedback, but reams of data delivered months after students leave school are as ineffective as the detailed criticisms written on the high school English paper mailed to the student weeks after final grades are assigned.

Let's take a closer look at how each of these FAST elements relates to feedback.

Fairness

My favorite lesson in fairness came from Mr. Freeman French, my junior high school orchestra conductor, who had students audition from behind a curtain. Neither students nor the teacher knew the gender, identity, ethnicity, or socioeconomic status of the player. We could only hear the music. While Mr. French's commitment to fairness may seem extreme, it represents a commitment to principle that seems elusive in the context of bias that ranges from Olympic skating to World Cup soccer in which, to put it mildly, fairness is not always the primary value on display. Certainly the blind audition approach of Mr. French had its limits—he ultimately had to look at his performers and give them feedback face to face, but the tone of fairness that he set in his classes conveyed the fact, as well as the impression, that our screeching strings—sharp and flat, too fast or

too slow—elicited his feedback solely based on our work and not our appearance.

I am certain that the vast majority of teachers aspire to be fair, but the distortions in feedback based on gender, ethnicity, and socioeconomic status are too consistent and too vast to be explained by performance alone. Similarly, the relationship between socioeconomic and ethnic variables and student test scores is consistent and pervasive (Putnam, 2015; Herrnstein & Murray, 1994), but those relationships say more about how popular tests value what wealthy white students know than they do about the intellectual incapacity of poverty-stricken minority students.

Unfortunately, the cure for unfairness in standardized tests can be worse than the disease, if teachers seek to compensate for unfairness by awarding high grades for poor performance. The pursuit of fairness at the expense of accuracy and specificity does not advance the cause of equity.

Accuracy

We would all like to think that our feedback is accurate. After all, people with college and graduate degrees know more than their students, so their feedback has to be accurate, doesn't it? In fact, it would be more precise to say that our advanced education has provided us with specific knowledge, while feedback must be an accurate reflection of what a student has learned or not yet learned. Factual accuracy and contextual accuracy are the first two principles of accurate feedback. The distinction between factual and contextual accuracy not only is important for classroom teachers but also for school leaders and policymakers. The third principle of accurate feedback is questioning. Asking students questions, rather than just making factually accurate statements, leads to a deeper understanding of students' thought processes.

Factual Accuracy

When teaching single-digit addition to first-grade students, the statement "No, Timmy, two plus three is not four; two plus three equals five" is factually accurate. However, before we take that for granted, it is worth recalling that many highly regarded tests fail

even this basic requirement. Almost every year, clever students find that there are either no correct answers or more than one correct answer on high-stakes college admissions examinations. Moreover, student writing is now required on the SAT, the Graduate Record Examinations (GRE), the Law School Admission Test (LSAT), and the Graduate Management Admission Test (GMAT), just to name a few exams with profound career and financial consequences for test takers. While great care is taken to provide consistency between two independent scorers or, in some cases, computerized scorers, this level of accuracy is short of being mathematical. According to a mathematical standard, two plus three equals five all the time, while the statement that an essay is worth 4.5 on a six-point scale is true only to the extent that other scorers agree with the judgment.

This is not meant to suggest that performance assessments, including work involving student writing, speaking, experimentation, and demonstration, are necessarily so subjective that they should not be used. Rather, an appropriate modesty about our professional practices requires that we check the accuracy of our feedback on performance assessments, just as we check the accuracy on a test of single-digit addition. While we cannot have perfect accuracy, we can have a level that is sufficient to lead students to improved performance.

This leads to the first principle of accurate feedback: a variety of observers, including other teachers, student peers, and the students themselves, must understand the criteria the teacher uses. In the previous example, how would you react if a teacher walked into the room and said, "Timmy, if you think that two plus three equals four, then that's fine with me, because the most important thing for me is that you love school!" Some readers would, of course, suggest that loving school and having mathematical accuracy are not incompatible. In fact, we would express concern that if students do not learn number operations through a steady diet of feedback, correction, and improved performance, the chance that they will love future schooling in mathematics is very low.

If we are going to require consistent feedback based on well-understood criteria, we should expect the same of every aspect of feedback, including those areas commonly thought to be subjective.

Students who receive feedback on a writing assessment know little about the criteria if they only can say, "I got a 2 on my paper." Another student, using the same scoring rubric, might be able to say: "I got a 2 on my paper because I left out the three supporting details I was supposed to put in the second paragraph. Also, my introduction didn't tell the reader very much about what I was going to write about. When my friend Laura read it, she said I should use more power words like *awful* or *horrible* instead of just writing *bad*. I think next time I can get a 3 if I work harder."

Note that this is not a problem of specificity—both students have the same rubric. Indeed, in many classrooms, rubrics overwhelm students. The problem is a fundamental issue of accuracy—if a student does not know what the rubric means, he or she cannot assess its accuracy. Just as a student can know that two plus three equals five and not four by engaging in experiments, so can a student understand that rich details, compelling introductions, and vivid language make for improved writing. Knowledge is not just the result of the teacher making pronouncements of judgment but of students understanding these judgments.

Contextual Accuracy

Let's return to the original feedback from Timmy's teacher. "Two plus three is not four," said the teacher, with unassailable accuracy. However, while no one would contest the next statement, "Remember, two plus three equals five," it does not necessarily meet the standard of contextual accuracy. Contextual accuracy requires that feedback reflect the context of what the student has and has not learned. In this example, how would we respond if the teacher said, "No, Timmy, two plus three is not four, and remember that standard deviation is the square root of the variance"? The expectation that first graders are able to grasp statistical functions is silly and irrelevant to the lesson. In the same way, we cannot provide feedback that accurately reflects what Timmy has learned and not yet learned until we do some further inquiry.

Questioning

Questioning as a means to provide accurate feedback is hardly new, as Socrates demonstrated more than two millennia ago. Sometimes it can lead a student to identify a thought process: "Why do you think that is true?" At other times, questions can help a student use another process to test his or her thinking: "What would the answer be if you used the blocks to find it rather than a pencil and paper?" Similarly, on writing and performance assessments, feedback that consists only of teacher statements (such as "Unclear sentence," "Awkward construction," "Improper usage," and "Another grammatical mistake") fails the standard of contextual accuracy.

For some students, these statements are as impenetrable as a gymnastics score during the Olympics. It's a mystery to me, but perhaps it's absolutely clear to an avid viewer of gymnastics competitions. Similarly, members of Garrison Keillor's mythical Professional Organization of English Majors (POEM) may be able to decipher what *awkward construction* means, but most of my students (including more than a few English majors) need some dialogue to sort it out. "Could you please help me better understand this sentence?" can help students distinguish clarity from opacity and replace an awkward expression with one that is more graceful.

While these examples apply to classroom interactions between students and teachers, the principles supporting accurate feedback apply in every context, including feedback provided to teachers and administrators. Whether in the form of annual test results, classroom observations, or formal evaluations, feedback does not meet the fundamental standard of accuracy if the criteria for evaluation are not understood clearly or if they are applied inconsistently.

When I discuss with teachers and administrators the value of feedback, they might respond with a weary, "We've already done that," referring to improved practices with rubrics, collaborative scoring, or student conferences. However, if I turn the tables and ask about their experiences when on the receiving end of feedback, they quickly dissect the inaccuracies and inconsistencies of the evaluations they received. "I'm the same teacher today as I was last year, yet depending on who is making the observation, I'm either superior or in need of

improvement. It's a crazy, unfair, and wildly inaccurate system!" So it is, and one of the central purposes of this book is to ensure that we provide students with feedback that is at least as accurate as we provide for the adults in the system.

Specificity

The third element of effective feedback is specificity. Feedback must be specific. There is a long-running dispute among assessment writers about the use of different kinds of rubrics and the relative merits of holistic and analytic rubrics. Grant Wiggins (1998) explains that a holistic rubric "yields a single score based on an overall impression" (p. 164). By contrast, an analytic rubric "isolates each major trait into a separate rubric with its own criteria" (p. 164). The essential principle, however, is not the label of the rubric but the degree to which it is applied consistently. Therefore, when we use rubrics that contain terms like *little evidence* followed by *some evidence*, *sufficient evidence*, and *superior evidence*, we are inviting chaotic ambiguity.

The antidote to ambiguity is not micromanagement. Students, administrators, and teachers have all uttered those words, signaling the surrender of independent thought ("Just tell me what to do") when ambiguous directions frustrate them. Hard work and collaborative efforts, when accompanied by failures in mindreading, can lead to disappointing results.

Certainly, hyperexplicit instruction is tempting. When we are specific with our requirements, students, teachers, and leaders may do exactly what is required, neither more nor less. Students produce essays with precisely five paragraphs, never more parsimoniously or extravagantly detailed than the formula suggests. Teachers print lesson plans precisely as prescribed and write state standards and learning objectives on the board. School administrators create strategic plans that conform to the format and style that higher authorities require. However, when feedback is too specific, students are assessed only on the extent to which they can follow a formula, not engage in learning. Telltale signs that well-intended rubrics have been subverted into mindless formulas include student work that almost uniformly begins with the same sentence.

How can teachers reach a balance between feedback that is too specific and feedback that is too ambiguous? Two principles, pulling in different directions, help teachers strive for this balance. The first involves boundaries and our interpretations of them, and the second involves consistency.

Boundaries

There are one hundred yards from one end of a football field to the other, not including the two end zones, and each team is allowed to place eleven players on the field, provided that the players remain on their own side of an imaginary line—the line of scrimmage. These boundaries are explicit and unchanging. The decision about where to place the players within them, however, is a judgment call. Similarly, fixed boundaries for classroom feedback include spelling and mathematical accuracy, but whether *egregious* should be substituted for *terrible* or whether bar charts should be oriented vertically or horizontally is a judgment call.

For five years, I served as the volunteer coach of a local high school debate team. Students have explicit boundaries, including time limits for speeches, the resolution that they must debate, and the integrity of the evidence they use during the competition. This experience serves as a useful illustration for the essential balance between explicit and flexible feedback on student performance. Applying excessively rigid boundaries for students may unintentionally lead them to become parrots instead of debaters, mimicking the arguments and evidence that they find on commercial websites that cater to affluent schools and students. Thus, the judges of these unfortunate students listen to the same speeches, arguments, and even grammatical errors, round after round, tournament after tournament. The students read the computer printouts, with their intensity interrupted only by the sound of the timer indicating that their time has expired.

If their competitors are locked into the same trap of specificity, then they might not refute their opponents' arguments but instead respond with another commercially prepared speech. These exchanges are not illuminating for the judges, competitors, or audience; therefore,

these debates represent the opposite of the critical thinking and rhetorical engagement that the activity might have offered to them.

An opposite but equally pernicious trend has emerged, which debate theorists label the critique (Hensley & Carlin, 2005). Students argue that debate resolutions are artificial boundaries and they should be free to address issues of greater importance to society. We will overlook, at least in this context, the fact that these students and their coaches embrace, without a hint of irony, rules about time limits, ballots, speaker points, and most especially, trophies. It is not that they truly want the absence of rules but rather they want to select the rules that are most familiar to them and least familiar to their opponent. It is as if one football team rolled onto the field equipped with tanks rather than shoulder pads but insisted that the time limits and boundaries of the playing field remain the same.

Where is the middle ground? In the classroom, as in other endeavors, there is a blend of specificity and creativity. Within a single debate resolution, students can think of hundreds of arguments. Within the rules of football, there are thousands of different offensive and defensive formations. One reason I enjoy watching cricket is the seemingly infinite variety of ways that players can respond to a unique combination of batsman and bowler. There is a balance between boundaries and freedom, and that is not only a key to effective feedback but to surviving in a free society.

Consistency

Expect consistency about boundaries but variations in student performance within them—a direct result of variation in judgment calls. How consistent is good enough? Feedback that is sufficiently specific should be consistent 80 percent or more of the time (Porter & Jelinek, 2011). Why is 80 percent important rather than 50 percent or 90 percent? When psychometricians, experts on measurement and testing in the field of psychology, evaluate high-stakes tests, they use the term *reliability* to assess the test's consistency. They expect students who answer a test item correctly to answer similar items correctly. Reliability is never perfect, but in general, the more consistent a test, the more reliable it is deemed to be. As a practical matter, testing experts generally expect about 80 percent consistency (a

reliability coefficient of 0.80) on tests for them to be deemed acceptably reliable; in other words, independent raters agree at least 80 percent of the time.

Classroom feedback should be similarly consistent. About four out of five teachers who look at a piece of student work should agree that it is proficient or not proficient. The more specific the feedback guidelines, the more consistent—the more reliable—the feedback.

Let's apply this theory in practical terms in the classroom. If five sources—three peers, a teacher, and a student doing a self-assessment—provide feedback on student work, four of those five sources should be consistent. That is, the feedback includes the same score on the rubric and very similar suggestions for improvement. This standard allows for some differences—interpersonal relationships can influence peer feedback, and some students can be hypercritical of their own work. However, if a student is not receiving consistent feedback, then the culprit may well be ambiguous qualifiers, such as *sometimes*, *sufficient*, *adequate*, or *good*, which invite different feedback on the same work.

Timeliness

When we say that feedback is timely, we mean that students receive the feedback with sufficient promptness to influence their performance. For example, in most sports, video games, and music rehearsals, feedback is virtually instantaneous—the bad play, wrong maneuver, or sour note is met with immediate feedback that leads students to stop and improve their performance. When conducting a science lab, solving a complex equation, or writing a paper, however, the length of time separating performance from feedback may be significantly longer.

In the context of student work, the extent to which we meet the standard of timeliness influences when the student can use feedback (Salem, 2013; Mester, 2011). For example, Lucy Calkins (1983, 1994), founder of the Teachers College Reading and Writing Project at Columbia University, pioneered a process in which students receive feedback from peers, teachers, and systematic self-assessment in order to improve the quality of their writing. Her rich illustrations

demonstrate the value of feedback (Calkins, 1983, 1994). The difference between poor writing and great writing (particularly from a gifted learner) is not the student's background or teacher but the quality, consistency, and frequency of feedback. I've watched Professor Calkins present these examples in small seminars and in large audiences of more than two thousand teachers, but despite changes in the learning environment, participating teachers' insight is invariably the same: they can't believe that the different writing examples are from the same students. The students are the same; what changes is the quality of the feedback.

The greater the number of students, the more challenging it is for teachers to provide timely feedback. Accomplishing this in secondary schools, where a teacher may have five different classes with thirty or more students each—more than one hundred fifty in all—is a particular challenge. However, it is not impossible. At Harlem Village Academy Charter School, more than 98 percent of students pass the New York Regents Exam, far above the city average (Harlem Village Academies, 2015). In this school, 100 percent of the student population is from a minority ethnic background and three-fourths are eligible for free or reduced-price lunch, a measure of low family income (U.S. News and World Report, 2015). Students routinely give one another feedback on everything from mathematics problem solving to behavior in the hallways.

Self-assessment also can be a rich source of timely feedback, provided that students are able to engage in an objective comparison of their work to a clear standard. An environment of consistent and timely feedback offers multiple benefits for a school, and it is the opposite of a system in which the only person qualified to provide feedback is the one holding the red pen, the gradebook, or the bullhorn. The essential requirement for timely feedback is that students have multiple sources of information, including a clear and consistent system of peer assessment and self-assessment. While it may take a teacher a day to grade homework and tests, student feedback is immediate, often within minutes of a peer's performance.

While the use of scoring rubrics since the 1990s has certainly been a positive step forward, there is wide variation in the quality,

consistency, and clarity of these instruments, particularly if students are to use them quickly and accurately. Larry Ainsworth and Jan Christinson's (1998) *Student-Generated Rubrics* is a useful beginning, but the ultimate test of rubric quality is not the words on the rubric itself but its impact on student work. For example, I have created what I thought were stunningly clear writing rubrics for my students, only to find that in their second drafts, they made the same errors—perhaps a bit more neatly—as they did in their first drafts. In other words, a well-crafted rubric met with students' uncomprehending compliance fails to deliver timely feedback. Therefore, while the expert who crafts a rubric may wish to include many elements, it is necessary to provide just enough feedback at the time it can be used to improve performance.

Effective athletic coaches are masterful at this, as are the most effective orchestral and choral conductors. While ineffective coaches hurl an unending stream of criticism and directions from the sidelines to their bewildered players, the best coaches give direction in a way that influences the action at precisely the right time. Similarly, it is interesting to watch how two of the great symphonic conductors of our time, James Levine and Lorin Maazel, move so little, even during rehearsals. Their feedback is precisely at the point in the music when it has the maximum impact. On the best athletic teams and in symphony orchestras, there is an exquisite network of communication—the alert from a fellow player, the arched bow, the lifted head—all of which reinforce and support feedback from the coach or conductor.

Good performance assessments, like good coaching, provide feedback that is FAST—fair, accurate, specific, and timely. The feedback is fair because it is consistent. Good rubrics provide the same score for the same performance. The feedback is accurate because it assesses precisely what it intends to assess. For example, an accurate assessment of writing considers the student's final draft after several attempts and clear feedback, not the first draft. Feedback is specific—consider the examples of the athletic coach and conductor—because students are able to respond to the feedback and demonstrate

improvement. Feedback is timely because, when students have a role in producing rubrics, they can assess their own work immediately and identify the next steps.

Table 3.1 summarizes techniques for ensuring fair, accurate, specific, and timely feedback.

Table 3.1: Elements of Effective Feedback

Elements of Effective Feedback	Techniques for Effective Feedback
Fair	Gender, ethnicity, socioeconomic status, or other characteristics of the students do not influence feedback. Teachers do not seek to compensate for biases in other tests by displaying reverse bias or awarding disadvantaged students higher grades for lower performances.
Accurate	Different observers, including other teachers, student peers, and the students themselves, understand the criteria the teacher uses to provide feedback. Teachers do not just make factually accurate statements to students; they ask students questions.
Specific	Boundaries are distinguished from judgment calls. Feedback on boundaries is consistent, with variations in expressions of student performance expected within those boundaries.
Timely	Feedback is delivered incrementally, at precisely the time when students can use it to improve performance.

In this chapter, we considered the four elements of effective feedback: fairness, accuracy, specificity, and timeliness. The next chapter turns its attention to the specific form of feedback known as grades, beginning with an overview of grading issues and the debate that surrounds them.