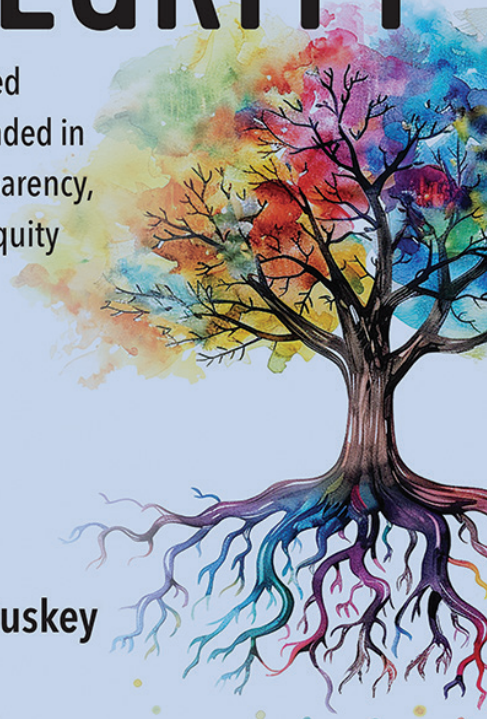


GRADING WITH INTEGRITY

A Research-Based
Approach Grounded in
Honesty, Transparency,
Accuracy, and Equity



Thomas R. Guskey
Nancy Frey
Douglas Fisher

CORWIN
Fisher & Frey

Thank you

FOR YOUR
INTEREST IN
CORWIN

Please enjoy this complimentary excerpt from
Grading with Integrity by Thomas R. Guskey.

LEARN MORE about this title!

CORWIN

Current Grading Schemas

What Are the Existing Problems?

As we discussed in Chapter 1, most current grading schemas have existed for at least 100 years. The problems associated with those schemas have also been recognized for most of that time. Although many reformers have recommended improvements, their recommendations have seldom taken hold. Why is change in grading policies and practices so hard to achieve? What makes grading reform so difficult? Some researchers suggest it could be a lack of understanding of the real issues that contribute to the intractability of existing grading problems. Other argue it is a lack of awareness of the research on grading—and, in turn, a reliance on the personal opinions and stories of consultants—that prevents implementation of more effective, evidence-based practices.

We chose the word *schema* intentionally in this context. In education and psychology, schema is a cognitive framework or concept that helps us to organize, interpret, and use information. Schemas allow us to interpret new situations based on our past experiences. Sometimes referred to as *scripts*, schemas are mental shortcuts that we use to figure out what to expect in a given situation and how to respond.

To show how this works, let's say you go to a fast-food restaurant in another country. Given your past experiences with this franchise in your own country, you expect certain things to be on the menu, like breakfast sandwiches, and you act in a way that is consistent with what you expect, such as skipping breakfast until you get there. What happens when you realize the franchise does not have breakfast sandwiches? It is likely this conflicts with your schema. If this is a one-time event, your schema probably won't change. You'll still expect the next fast-food restaurant you visit to serve breakfast sandwiches.

But if you keep encountering this situation, your schema will be updated to include the idea that not all fast-food franchises have breakfast sandwiches. At that point, you may start eating breakfast at your hotel before you go out for the day. Without schema, we would have to relearn and reevaluate our actions with each new experience, irrespective of whether or not we have had a similar experience in the past.

.....
If we fail to confront the problems associated with our current schemas, we will not revise and update them.

 Because of our past experiences in school, we all have schemas regarding grading policies and practices. Our experiences have shaped our thinking about grading, its function and purpose, and what it means for students and teachers. If we fail to confront the problems associated with our current schemas, we will not revise and update them.

Jean Piaget, the psychologist first to identify schemas, noted that adaptation is the process we use to update our schemas to match the world around us. Adaptation occurs in two ways:

- *Assimilation*: Solving new experiences using existing schemas
- *Accommodation*: Changing existing schemas to solve new experiences

Connect and Conclude

What is your grading schema? What are the factors that influenced the way that you think about grading?

Both assimilation and accommodation involve new experiences. One way to ensure schema adaptation happens is to experience failure. But because the stakes are high and involve the well-being of students, we certainly cannot force educators to directly experience failure in order to revise their grading schema.

What we can do, however, is present failures in grading systems that resonate with them in such a way that it opens their schema to adaptation.

In this chapter, we discuss four dimensions that inform a grading with integrity framework by exposing flaws in current grading schemes. We do this by examining four questions that drive the conversations and the controversies about grading. Each of these has the potential to add to or diminish the integrity of grading:

- What is honest?
- What is transparent?
- What is accurate?
- What is equitable?

WHAT IS HONEST IN GRADING?

Too many parents/families have heard their child say, “My teacher gave me a C because she doesn’t like me.” Right or wrong, the fact that a young person can raise an argument to support this claim illuminates a fundamental dilemma: students’ perceptions about the lack of honesty and fairness in grading are substantial. In fact, students’ perceptions of grading and reporting are tied to their justice judgments, which are their beliefs about personal responsibility, deservedness, and distribution of rewards (Dalbert et al., 2007). And therein lies the rub: grades are commonly viewed by students as a reward or punishment, rather than as an honest and objective evaluation of their performance. The inconsistency of grading practices from one teacher to the next also contributes to this sense of unfairness.

The grading schemas used by teachers strongly impact students’ perceptions of honesty and fairness. A study by Claudia Dalbert and colleagues (2007) involving students in grades 7–12 illuminates the interaction between how grades are distributed and students’ justice judgments. In this study, students were presented with four scenarios and asked to rate the justness of different grading schemes. Here’s an example:

A talented student generates a proof in a math test, but the solution is wrong. What would be a just grade to give them?

In each case, students were presented with three grading schemas: norm-referenced, criterion-referenced, or self-referenced, meaning the grade compared students’ current performance to their past performance. Students overwhelmingly identified criterion-referenced grading decisions as being honest and highly just and individually referenced grades as being somewhat just. By a large margin, they ranked norm-referenced grading as being the most unjust (Dalbert et al., 2007).

Teachers sometimes have difficulty in aligning judgments about honesty and fairness to their grading practices. A well-known study of secondary teachers’ grading practices by Rick Stiggins and colleagues (1989) found that teachers applied less than half the recommendations of researchers offered in educational measurement textbooks.

Grading With Integrity

Robin Tierney and colleagues (2011) interviewed 77 tenth-grade mathematics teachers regarding their beliefs about grading in light of their province's newly adopted standards-based (i.e., criterion-referenced) grading policies. Results revealed that 75 percent of teachers said that they still considered the students' trajectory of improvement (i.e., self-referenced) and the amount of effort students exerted when calculating grades—two practices that are not to be a part of the province's grading framework. Furthermore, many of the teachers cited "professional judgment" as an umbrella explanation for deviating from the province's standards-based grading policy.

Another study of middle school teachers by Aimee Howley and colleagues (2000) also found troubling subjectivity imposed in teachers' judgments from one student to the next, "confounding effort and compliance with achievement" in the construction of individual student grades (p. 232). When confronted with the disparity between a student's high scores on a standards-based assessment and her low grade point average, one teacher in the study said,

I don't know why she's in that top group. . . . The top group are kids who would be on your honor society, which we recognize for good citizenship, model students, A and B students. Clarissa would not be there, because a lot of teachers would not recommend her because of her attitude. (p. 229)

The researchers noted that the variance in grading practices among teachers seemed to be tied to what they termed "an ethos of effort," a condition of the classroom climate that prized control over achievement (p. 229). Interestingly, they reported that there was also an unspoken bargain afoot: in return for compliance in these classrooms, the teachers made the tasks easier for students.

In some cases, the professional judgments of the high school math teachers in Tierney's study were undoubtedly appropriate. But when teachers' grading decisions are divorced from research-based principles and are further muddied with non-achievement indicators like compliance, responsibility, and effort, the stated purpose of grading crumbles. No wonder so many students perceive grading as unfair. Tierney and colleagues concluded that "without a better understanding of the most essential grading principles, teachers may continue to struggle with assessment policies in a standards-based system" (2011, p. 223).

In the absence of clear grading criteria, honesty and fairness are undermined. Jeffrey Schinske and Kimberly Tanner (2014) conducted a review of grading practices that demonstrated the tremendous variability in those practices among teachers. This variation was influenced not only by the student's identity but also by factors such as the quality of the handwriting (Bull & Stevens, 1979) and the order in which papers were reviewed (Spear, 1997).

Teachers' judgments and the accompanying grades they assign have also been shown to be influenced by a phenomenon termed the "halo effect." This is the tendency of allowing students' previous performance to influence judgments of their current performance. For example, suppose your favorite actor starred in one film that was terrific, but that same actor's next film was terrible. The halo effect might influence your rating of the new film, giving it higher marks than it deserved because of your favorable opinion of the actor.

In a study by Margaret Spear (1996), teachers rated middle school science essays differently depending on whether the student expressed a high or low level of interest in the field of science on a different writing task. The essays of students who expressed high levels of interest in science were scored higher. When the same essay was paired with an expression of little interest in science, it was scored lower, even though the content of the essay was identical. The halo effect, or in this case, the "horn effect," biases teachers' grading.

We point out these flaws to highlight the fact that grades will always be subjective to some degree. The act of grading involves one group of human beings (i.e., teachers) making judgments about the performance of another group of human beings (i.e., students) and then often communicating those judgments to yet a third group of human beings (i.e., parents and families). The judgments of any of these groups are naturally subjective and can be fallible. The challenge of grading with integrity is to limit and hopefully eliminate the unconscious biases that often influence those subjective judgments. We must ensure the honesty and fairness of those judgments if we are to meaningfully communicate valid information about students' learning progress and improve learning success. In later chapters, we offer specific, practical steps educators can take to do just that.

.....
The challenge of grading with integrity is to limit and hopefully eliminate the unconscious biases that often influence subjective judgments.

Summarize and Speculate

How do we make grades more honest and accurate? What can be done to confront the various forms of bias that creep into grading systems?

NUMBERS AND GRADES

Numbers are integral to all forms of scientific judgment. Quantitative evidence serves as the foundation of scientific knowledge by providing the basis for testing scientific theories and hypotheses. Lord Kelvin's assertion, "If you cannot measure it, you cannot improve it," underscores the significance of sound measurement in scientific pursuits. However, gathering quantitative evidence on human subjects is considerably more complex than it is with other scientific phenomena, primarily because it involves both direct and indirect approaches to measurement.

Scientifically, **direct measurement** involves explicitly measuring the characteristic of an object or a person that the researcher aims to quantify. For instance, consider the measurement of a student's height (Figure 4.1). In this process, the student would stand with their back against a wall, a level instrument like a ruler or book would be placed on the top of their head to mark the wall, and the distance from the floor to that mark would be measured using a device like a measuring tape or yardstick. The recorded number represents the direct measurement of the student's height in inches or centimeters.

Most of the measures teachers use to determine students' grades, however, are not direct measures—they are indirect measures.

Figure 4.1 • Example of Direct Measurement



Scientifically, **indirect measurement** involves measuring something else and converting it into a measurement of the characteristic in question. In education, for example, teachers cannot directly measure students' achievement or proficiency by placing some measuring device upon them. Instead, teachers ask students to answer a series of questions or perform certain tasks. Teachers then make *judgments* or *inferences* about students' level of achievement based on their responses or performance. Because these judgments or inferences are subject to personal interpretation, indirect measures are more susceptible to bias and interpretation errors compared to direct measures.

Upon learning of this distinction, some teachers respond, "But wait. The percentage of questions students answer correctly is a direct measure of their achievement, isn't it? Doesn't 80 percent correct mean a student has mastered 80 percent of the content or learning goals?" However, interpreting something as seemingly simple as "percentage correct" is far more nuanced than most teachers imagine. For instance, interpretations can vary depending on the format of the questions, the difficulty and complexity of the questions, the alignment of the questions to the instruction, the time students were given to respond to the questions, and so on. The accuracy of percentage correct and percentage grades is far more illusionary than real (Guskey, 2013).

.....
 The accuracy
 of percentage
 correct and
 percentage
 grades is far
 more illusionary
 than real.

Failure to recognize the difference between direct and indirect measures often leads to false assumptions about the numbers assigned to students, especially when considering the validity of a score. This is called the *illusion of data validity*, and it leads to the false belief that the information we collect from students is always honest, complete, and accurate (Jansen et al., 2022). This is rarely true when it comes to tallying the results from indirect measures of student achievement for the purpose of grading.

There are two reasons that the indirect measures of student learning we use in determining students' grades are generally less precise than direct measures of students' height or weight. First, working with people introduces a host of variables that can confound our grading scheme. For instance, one student may be feeling anxious about the deadline for the essay and rushes their final draft. Another may have a low-level fever, and a third might have had an argument with a parent that morning before school. Any of these situations can impact students' performance on a given day—and, in turn, the judgments we make about their performance.

Grading With Integrity

Second, most of us are not skilled psychometricians with advanced training in educational measurement. We rely on a collection of evidence gathered from teacher-made assessments, commercial curriculum item banks, and, increasingly, AI-generated tests. These are not necessarily bad, but each has inherent flaws. These imperfections lead to potential errors in our judgments about the meaning of students' scores. The real danger, however, lies in giving in to the illusion of data validity. Just because a particular assessment yields a number calculated in a mathematically precise way does not mean that the inferences made and the conclusions drawn about that number regarding students' performance are accurate or complete.

According to the 2019 High School Transcript Study, the most common grading scheme used in schools are letter grades that are converted from a 100-point percentage scale (National Assessment of Educational Progress, 2022). However, these grade conversion schemes vary widely from school to school. In some schools, for example, the range for an A is 90–100, while in others it is 93–100. In some schools, teachers add + and – to the letter grades, and in other schools they don't. The one common element in nearly all schools is that anything below 60 percent correct is considered a failing grade that is usually labeled as an F in the grading scheme. Nevertheless, a major problem concerns the reliability and defensibility of the score itself. For example, would all teachers in the same school similarly articulate and defend the difference between a 59 and a 61? How about the difference between an 89 and a 91?

Deriving letter grades from the 100-point percentage grading scale is further tilted against students. If anything below a 60 is considered an F, signaling impending or actual failure, then the percentage grading scale consists of 60 “distinct levels of failure and only 40 levels of success” (Guskey, 2013, p. 70). In other words, nearly two-thirds of the scale scores designate failure! If we accept the fact that our measures of student learning are estimates (educated ones, but still estimates, given that most are indirect measures), then we must also accept the truth that a 100-point grading scheme is too fine-grained for reliable or consistent use. We must also accept the imbalance of devoting the majority of the scores in the scale to chronicling student failure.

The levels of performance required in most grading schemes exceeds what we can consistently and reliably determine. Again, consider the 100-point grading scheme. Acceptance of this grading scheme means that there are 101 levels of performance to be gauged by the teacher. (Keep in mind that 0 is a

level of performance.) Even the letter-grade scale with + and – modifiers leaves 13 discrete levels of performance teachers need to justify.

More school systems are taking action to further reduce the levels of performance by gearing their grading through the use of simplified letter or standards-based grading schemes. And let's not forget what many state Departments of Education, as well as many colleges and universities, did in response to pandemic schooling: they reduced the levels of performance to just two: Pass/No Pass or Satisfactory/Incomplete. We will discuss these alternatives at length in Chapter 5. Figure 4.2 shows a comparison of grading schemes according to levels of performance.

Connect and Conclude

How are numbers and grades integrated in your current grading system? What indirect measures of student achievement are used? Is there an assumption that these indirect measures are direct measures or is the illusion of data validity recognized in the grading systems used in your school?

Figure 4.2 ♦ Grading Systems and Number of Grade Categories

GRADING SYSTEM	LABELS	LEVELS OF PERFORMANCE
Percentage Grades	100, 99, . . . 2, 1, 0	101
+ and – Letter Grades	A+, A, A–, B+, B, B– . . . D, D–, F	13
Letter Grades	A, B, C, D, F	5
Standards Based	Exemplary, Proficient, Not Yet	3
Satisfactory/Incomplete	S, I	2

WHAT IS TRANSPARENT GRADING?

Transparency in grading implies that the grades assigned are true to the stated purpose and that the criteria for assigning grades are clear and explicit. Unfortunately, rather than being an indicator of student achievement or proficiency, grades are often viewed as a tool for motivating and controlling students. In 1831, Harvard’s president, Josiah Quincy, wrote that grades are “the best assurance for the continued and unremitting attention of students to their exercises,” reflecting a belief that has percolated for centuries (cited in Schneider & Hutt, 2014, p. 206). In this context, it is useful to first consider what actually prompts motivation in humans—and specifically in students.

Motivation is often described through the lens of self-determination theory (Ryan & Deci, 2000). This theory posits that motivation moves from strictly extrinsic systems (e.g., rewards, points, or other benefits) to more intrinsic ones (e.g., increased pride, satisfaction, and confidence) when three important needs are met: *autonomy* (a sense of agency and influence over our life), *competence* (a belief in our abilities), and *relatedness* (the relationships we have with others). In the context of performance, autonomy has been identified as the most powerful, as students with a higher degree of autonomy are more likely to exert the necessary energy to achieve goals (Niemi & Ryan, 2009). In contrast, grades in isolation do not seem to be a motivator for learning and can fuel the desire to avoid challenge (Chamberlin et al., 2023). Nevertheless, when grades are paired with feedback about competence, motivation increases (Koenka et al., 2021).

.....
**Transparency in
 grading implies
 that the grades
 assigned are true
 to the stated
 purpose and
 that the criteria
 for assigning
 grades are clear
 and explicit.**

In addition to the belief that grades are motivating, many educators, policymakers, and the public believe grades serve to encourage timely performance. It is quite common for teachers to lower grades or assign a score of 0 for assignments that are submitted late in an effort to force students to “face the music” (Tyner & Petrilli, 2018, p. 16). Their rationale is that there are time-based expectations in “the real world”—a misnomer that communicates that school is somehow imaginary. We have heard examples such as filing a tax return, paying bills, and meeting deadlines employers pose for projects. “Their boss isn’t going to give them an extension on a project” is generally how the thinking goes. There are also many

examples of missed deadlines that come with financial consequences. If you're not at the airport on time, the plane is going to leave without you, and there won't be a refund.

But the actual leeway in the adult world is broader than we like to tell young people. How many teachers have failed to submit their timecard by the due date? It doesn't result in receiving no pay for work completed, although it may result in a delay of payment. There is also little evidence about uncontested deadlines in the workplace. A series of four studies by Ashley Whillans and colleagues (2022) examined the impact of requesting deadline extensions in professional workplaces and in college. The researchers made the following discoveries:

- Supervisors reported that 53 percent of deadlines were actually flexible.
- In contrast to employee concerns, supervisors did not rate workers who requested extensions more negatively.
- Supervisors and teachers rated the quality of the work completed as higher as a result of the extension.
- Students and employees incorrectly believed their instructor's or employer's perceived competence of them would be damaged, and therefore they were reluctant to ask for an extension to the deadline.

The researchers stated,

Deadline extensions can improve performance quality at the expense of speed. . . . Our data show that when evaluation criteria are ambiguous and quality is subjectively measured . . . employees focus more on impressing supervisors by optimizing speed over quality. . . . [The] focus on performance speed is consistent with research showing that people tend to maximize easily measured mediums, such as *points* and *time*. . . . Yet, in some contexts—such as when the importance of quality is explicitly stated—quality might be prioritized over speed. (p. 10)

.....
The integrity of grading is diminished when the rationale for using grades is to motivate students or to teach character development.

These findings further reinforce the notion that our sense of competence is influenced by concerns about how others will perceive us. Researchers call this “self-presentation.” Concerns about self-presentation contributed to employees' reluctance

to ask for an extension, even though the quality of their work would be improved. In addition, this hesitancy to seek help, an important expression of autonomy, was more apparent in tasks where success criteria were vague. In these cases, time and points earned—not quality—took center stage.

Summarize and Speculate

What role does grading play in motivating students? Are grades in and of themselves a source of motivation? Do grades impact timely performance on tasks?

In sum, the usefulness of grades as motivation for excellence and as a means to teach non-achievement behaviors such as responsibility and time management is dubious. Grades have a purpose, as we noted in Chapter 3. Transparency in grading ensures that the grades assigned are aligned with that purpose and that the criteria used to determine grades are explicit and clear to all interested parties. That transparency is lost and the integrity of grading is diminished when the rationale for using grades is to motivate students or to teach character development.

WHAT IS ACCURATE GRADING?

A third dimension of grading is that it should communicate accurate information efficiently and effectively. However, a one-size-fits-all approach to grading is problematic when we consider who the various stakeholders are. One may be a seven-year-old student, while another stakeholder is the seven-year-old's family. A third may be the school system itself. And let's not forget future interested parties, such as that student's middle school and high school teachers. Why would we possibly think that one mechanism alone could accomplish all these diverse communication needs?

The reality, of course, is that it can't. In an effort to meet the needs of these diverse groups, we use computerized grading programs that include gradebooks, report cards, and permanent records or transcripts. However, the purposes of these different reporting tools often get mixed up. To gain additional clarity, it is helpful to consider the role of each of these reporting tools.

A gradebook is an *ongoing record of performance* for students and their parents/families, while a report card is an *interim*

summary of performance for students, parents/families, and immediate educator colleagues. These are different from permanent records and transcripts, which are *summary judgments of performance* for students and parents/families, as well as the educational system and third parties (see Figure 4.3).

Figure 4.3 ♦ Components of Computerized Grading Programs

COMPONENT	GRADE BOOK	REPORT CARD	PERMANENT RECORD/ TRANSCRIPT
What does it include?	Scores	Grades	Summary grades
Purpose?	<i>Ongoing</i> record of performance	<i>Interim</i> summary of performance	<i>Summary judgments</i> of performance
Who has access?	Parents, families, and students	Parents, families, and students	Parents, families, students, and third parties

Source: Guskey, T. R. (2024). *Engaging parents and families in grading reforms* (p. 39). Corwin.

The ready availability and ease-of-use of computerized grading programs gives them a seductive quality that can lull teachers into false beliefs about the accuracy of grades due to the mathematical precision of the score tabulations used in determining grades (Guskey, 2024). Much of the confusion lies in the differences between what a gradebook offers and what a report card provides. It is essential that school leaders communicate these differences to all stakeholders, and especially to teachers and parents/families. Perhaps the single most crucial message is this: “Not everything recorded in the gradebook is used in determining students’ academic report card grades” (Guskey, 2024, p. 41). While nonachievement factors such as homework completion, formative assessment scores, and attendance may be recorded in the gradebook, these data shouldn’t be used as measures of academic performance. Rather, these are learning enablers that should be recorded separately on report cards and on the transcript, apart from students’ academic grades. We’ll discuss reporting nonachievement learning enablers and other factors in Chapter 6.

.....
Nonachievement factors such as homework completion, formative assessment scores, and attendance shouldn’t be used as measures of academic performance.

WHAT IS EQUITABLE GRADING?

Reflect and React

What is the purpose for each of the components of the computerized grading programs used in your organization? On a scale of one to four, with four being highly accurate, how would you rate each of them (e.g., gradebooks, report cards, permanent records, and transcripts)?

At the most basic level, equitable grading “aims to measure how students understand the classroom material by the end of a term without penalties for behavior” (Randazzo, 2023). Joe Feldman (2023) argues that equitable versus inequitable grading centers around three issues:

1. The variability of grading practices
2. Mathematically unsound calculations
3. Biases

Each of these elements has a negative impact on the accuracy and equitability of students’ grades.

But equity involves more than overcoming these three hurdles. It requires us to recognize that students do not start at the same place or learn at the same rate. The first challenge—not starting at the same place—can be the result of a lack of opportunities to learn due primarily to differences in students’ social, demographic, or economic backgrounds. These are factors over which students have no control. Thus, their grades may be influenced by the cumulative experiences, trauma, discrimination, ableism, sexism, homophobia, and racism that has impacted their world.

In some places, “equity grading” is interpreted to mean simply that students’ grades are not affected by missing homework assignments or turning in work late or not at all. To our thinking, however, this is wrong-headed. As we will emphasize in Chapter 6, these aspects of students’ behavior should be reported on both the report card *and* the transcript. However, they *must be reported separately*, and they should not be included in determining students’ academic achievement grades. In this way, students and their parents/families receive crucial feedback on these aspects of performance, but these pieces are reported separately from an academic grade that represents how well students have learned and what they are able to do.

A group of district leaders and university faculty in San Diego met over the course of several years to improve the outcomes of students and ensure equity for all. As part of their work,

they developed a definition of “equity in education,” which reads as follows:

- Equity recognizes that every student comes to school with a unique identity profile that is too often impacted by racism, bias, or bigotry.
- Equity occurs as a result of sensitive, courageous, and creative conversations and actions.
- Equity requires the distribution and redistribution of resources and initiatives based on individual and group needs derived from multiple sources of qualitative and quantitative data. Equity leads to engaged, inspired, and successful learners. (Fisher et al., 2019, p. 45)

Although this statement was not designed specifically around grading, it has important implications for how teachers grade. First, it compels teachers to recognize students’ unique histories and identity profiles. Furthermore, it encourages educators to engage in meaningful conversations that focus on data allowing for the equitable distribution and redistribution of resources and programs. This is a critically important point in discussions regarding equity in grading. Based on the qualitative and quantitative data instructors collect, what needs to be done to ensure adherence to the promise of equity?

The pursuit of equitable grading involves much more than simply allowing students to have multiple tries on assessments. It requires that the needs of students are systematically addressed and that actions are taken with regard to assessment data and other grading information that might involve implicit and disguised inequities. For example, such actions might include teacher-guided corrective instruction or collaborative peer tutoring prior to retakes on assessments. They might also include examining the way knowledge is represented on assessments or detailed explorations of inherent grading biases.

Summarize and Speculate

How equitable are the grading practices in your school? How do they recognize that students do not start at the same place or learn at the same rate?

BIAS IN GRADING

The issue of grading bias takes two forms. The first is in relation to the design and interpretation of assessments. Earlier in this

Grading With Integrity

chapter we described how design and interpretation issues can contribute to the illusion of data validity. The second form of grading bias concerns relational biases stemming from teacher knowledge of students' backgrounds. Each of these forms of bias are threats to the integrity of grading.

Relational biases are deeply entrenched, and they can play a strong and compelling role in how teachers grade. Studies have documented the effects on grading of a variety of student characteristics, including students' weight (Finn et al., 2020), gender (Di Liberto et al., 2022), socioeconomic status (Doyle et al., 2023), home life (Hardré & Sullivan, 2008), and race (Woods Jr., 2023). A meta-analysis of studies on grading bias by John Malouff and Einar Thorsteinsson (2016) analyzed the results of twenty studies involving 1,935 graders of students' work from primary grades to college. The hypothesized biasing characteristics included different race/ethnic backgrounds, education-related deficiencies, physical unattractiveness, and poor quality of prior performance. They found that bias consistently occurred across all levels when graders were exposed to these specific types of information about students other than students' performance on a task.

While we often think of bias as having a negative impact on grading, Malouff and Thorsteinsson (2016) found it sometimes works in the opposite direction, with some students receiving a positive bump in their grade due to a "reverse bias." Much of this bias is unconscious, and most studies are able to determine the impact but not the source.

To be sure, many teachers take steps to reduce implicit or unconscious biases in their grading practices, such as masking the names of students, asking for the opinions of others, and devoting adequate time to grading. Important research also shows that implicit bias is more likely to influence decisions when there is incomplete information or when people are feeling overloaded and pressured (Bertrand et al., 2005; Johnson et al., 2016).

PUTTING ISSUES OF FAIRNESS AND EQUITY TO THE TEST

Let's bring several concepts together as they relate to the dimensions of honesty, transparency, accuracy, and equity. First, we need to accept the limitations of interpreting

evidence, especially what is gathered from indirect measures, because no collection of evidence ever provides a complete picture of a student's performance or achievement. Second, we must recognize that although implicit biases are present and not fully under our conscious control, we can take specific steps to mitigate them.

A study by David Quinn (2020) titled "Experimental Evidence on Teachers' Racial Bias in Student Evaluation: The Role of Grading Scales" directly addresses these concepts and was named one of Edutopia's ten most influential studies of the year. Quinn tested a theory of racial bias in grading through a web-based study of 1,549 K–12 teachers, who were asked to evaluate a writing sample of a second-grade student. There were two variables at play. The first had to do with race. While the writing samples were identical, the names used were race signaling. The writing samples are featured in Figure 4.4a and 4.4b.

Participants were then asked to grade the writing samples using one of two grading criteria: vague or specific. The vague grading criteria did not include performance indicators but only a seven-point Likert scale ranging from "far below grade level" to "far above grade level." These scales were converted to a binary "at or above grade level" or "below grade level" to facilitate calculation. The specific grading criteria were shown in the form of a rubric about recounting details in a narrative event.

When teachers evaluated the two writing samples using the vague grading scale, the "Dashawn" paper was scored nearly 5 percentage points lower than the "Connor" paper. However, when the same teachers scored using the specific grading criteria, there was no difference in the results. As Quinn (2020) noted,

Teachers' stereotypes may have more influence on their evaluations when they are not given clear, specific criteria on which to rate student work. In contrast, teachers may be less likely to draw on their stereotypes when they have less discretion over the criteria for evaluating students. (p. 388)

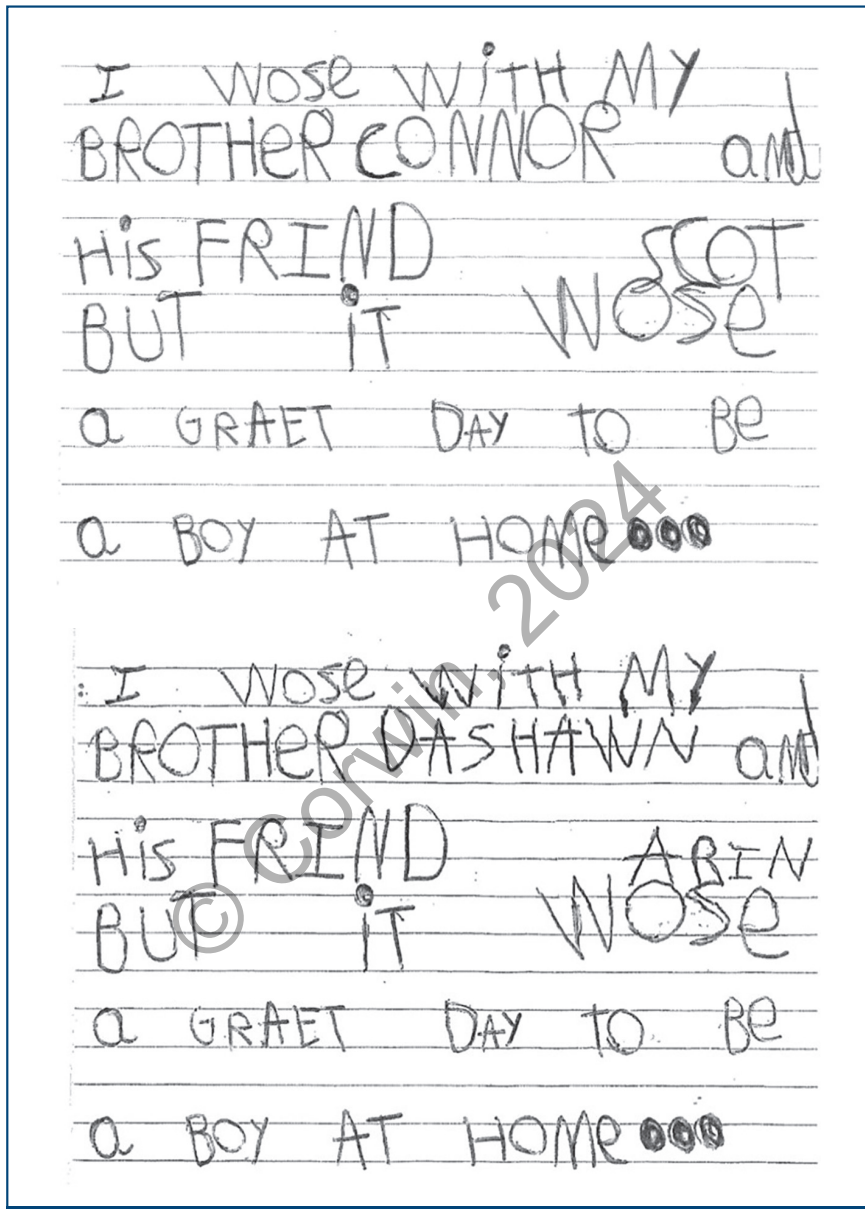
In this case, increasing the fairness of the scoring and grading through the use of specific performance criteria reduced the inherent inequities that would otherwise surface.

.....
Although implicit biases are present and not fully under our conscious control, we can take specific steps to mitigate them.

Reflect and React

Should student performance used for grades be evaluated anonymously?
 What are the pros and cons of grading work without student identifiers?
 What are the challenges in doing so?

Figure 4.4a and 4.4b • Writing samples



Source: Quinn (2020, p. 392)

By “engineering evaluation procedures in bias-minimizing ways,” the researcher demonstrated that it is possible to eliminate the effects of such unconscious biases (p. 389). The adjustment of the scales used for evaluations can provide educators with a tool for mitigating biases while forwarding more equitable grading policies.

MINIMUM-GRADE POLICIES DO NOT YIELD EQUITY

“Equitable grading” has become a popular term in the education literature, especially since the COVID pandemic. Yet, despite its popularity, the issues involved in implementing truly equitable grading practices are far more complex and require more substantial change than most educators anticipate. As a result, efforts to implement more equitable grading practices are often thwarted by significant parent opposition, especially when critical dimensions are overlooked (see Coffey, 2023; Habeeb, 2023).

The debate over “minimum-grade” policies offers an excellent example. These policies have become increasingly popular in schools and school districts that require teachers to assign percentage grades ranging from 0 to 100 but that also want to implement more equitable grading practices. To alleviate the devastating effects of assigning a score of zero, which generally dooms a student to failure because it is such an extreme score in the percentage grading scale, minimum-grade policies prohibit teachers from assigning a percentage grade lower than 50 percent to any assignment or assessment, regardless of the student’s actual performance. Although well-intentioned, these policies tend to be misplaced efforts to rectify grading schemes that are inherently problematic, resulting in a Band-Aid approach that dresses a wound without healing it. Minimum-grade policies address a problematic symptom but neglect the cause.

Even when teachers assign a minimum percentage grade of 50 percent, they still have the dilemma of a grading scale with 51 discrete levels of performance for which it is impossible to gain an acceptable level of consistency or reliability in teachers’ judgments. Plus, significant numbers of teachers and community members strongly object to the idea of assigning students half credit (i.e., 50 percent), even if they do nothing. That is why well-reasoned arguments based on mathematical scale differences rarely convince sceptical teachers, parents, board members, or legislators of the merits of minimum-grade policies, and they sometimes result in legal challenges that outlaw the use of minimum grades (Rapoport, 2010; Robelen, 2010).

Furthermore, focusing narrowly on minimum-grade policies that prohibit zeros while failing to address subjectivity and

.....
Minimum-grade policies address a problematic symptom but neglect the cause.

Grading With Integrity

bias in grading is likely to spark increased resentment among educators, parents and families, students, and other stakeholders. In many situations, these issues are further exacerbated by a lack of professional development about effective assessment practices,

A more winnable and far more effective approach in schools or school districts that require percentage grades is to establish a purpose statement that indicates report card grades should reflect “students’ current level of performance” or “achievement at this time.” This purpose statement then becomes a basis for challenging the practice of averaging scores over time to determine students’ grades. Like minimum-grade policies, this approach negates the influence of the occasional extremely low score or zero. But because it aligns with an agreed-upon purpose statement, it is generally more acceptable to a broader range of stakeholders.

Minimum-grade policies are actually an overcorrection for grading schemes that do not make it any more likely that students will be successful. In addition, they defeat the communicative intentions of the grades themselves. Is it any wonder that teachers, parents, families, community members, and legislators are baffled by a policy that awards half credit for nothing? Or that students misinterpret the practice as meaning that they don’t need to do much of anything to demonstrate their competency? Instead, why not address the real problem, which is that current procedures used to determine grades need to be carefully examined and revised to ensure the grades assigned are indeed honest, transparent, accurate, and equitable.

SUMMARY

There are several problematic features of different grading schemes, including issues related to fairness and equity. Unfortunately, many grading schemes fall victim to entanglements around usefulness and, as a result, fail to provide students with accurate information about their current performance. When we confront these failures, we are more likely to update our schemas and develop and implement more effective schemes for grading. These new schemas need to center on integrity and ensure that the information shared is honest, transparent, accurate, and equitable.

To strengthen grading policies and practices—and to truly grade with integrity—we must confront barriers to fairness and usefulness. In addition, we must meet the needs of all educational stakeholders regarding communication, and we need to do so in ways that are truly equitable. Attempts to quickly solve serious problems (e.g., the effects of assigning zeroes in percentage-grading systems, averaging, computerized scoring, etc.) with Band-Aid approaches like minimum-grade policies are likely to have unintended negative consequences, and these consequences can make more effective, comprehensive reforms impossible. To borrow a sentiment from the medical profession, “We can’t say, ‘The surgery was successful, but the patient died.’” To bring integrity to grading, we must know the pitfalls of different grading schemes and work purposefully to counter those drawbacks in order to build a stronger system that serves students, educators, and parents and families well.

Interpret and Infer

1. Reflect on the questions posed in this chapter: Is our current grading scheme *honest*? Is it *transparent*? Is it *accurate*? Is it *equitable*?
2. How is your schema about grading being updated? What information did you assimilate, and what information did you accommodate?
3. How can you reduce bias in grading? Of course, the first step is to recognize that grades are susceptible to bias, but what then?