

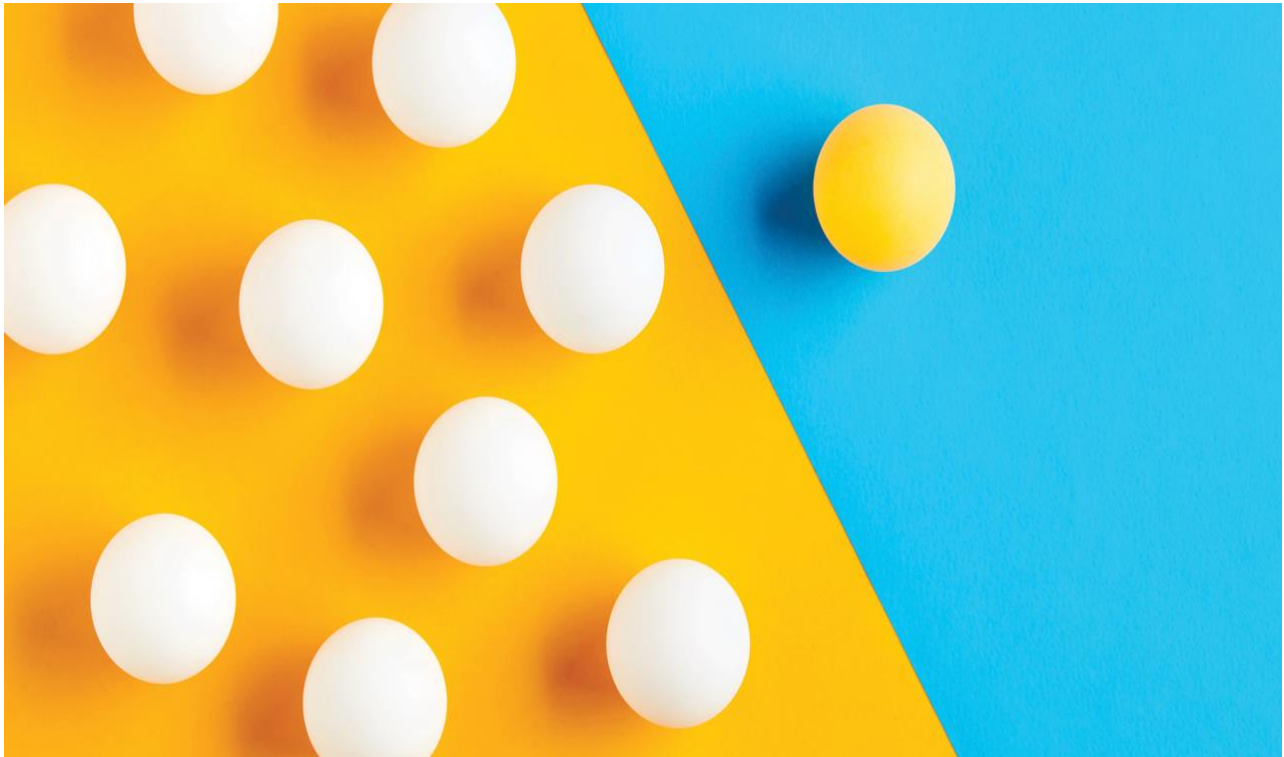
## Addressing inconsistencies in grading practices

Thomas R. Guskey

*Phi Delta Kappan*, 105(8), 52-57.

Apr 29, 2024|

<https://kappanonline.org/addressing-inconsistencies-in-grading-practices/>



***Coming to agreement about the purpose of grading and establishing clearer and more accurate reporting structures can pave the way for more learning-focused grading systems.***

Throughout the world today, school leaders are struggling to implement grading reforms. They recognize that many current grading policies and practices are outdated and inadequate. They also know these policies and practices don't align well with recent changes in school curricula, instructional strategies, and procedures for assessing student learning. Yet despite their commitment and good intentions, these dedicated school leaders are facing unanticipated opposition.

Grading reform means challenging some of education’s longest held and most firmly entrenched traditions (Guskey & Brookhart, 2019). These challenges prompt concern among all stakeholders and serious opposition from some. In many cases, the most adamant opposition comes from parents and families, especially for reforms involving standards-based or competency-based grading (Franklin, Buckmiller, & Kruse, 2016; Young, 2023).

## AT A GLANCE



- School leaders struggle in their efforts to implement standards-based grading reforms in large part because they fail to address the inconsistency in grading practices among individual teachers.
- Thomas R. Guskey recommends three steps to improve consistency and lay the foundation for grading reform:
  - Reaching consensus on a purpose statement for grading.
  - Using grading scales with four to seven categories of student performance
  - Reporting separate grades for academic achievement and non-academic aspects of students' performance.

### Sources of frustration

Ironically, few parents and families oppose the basic principles of standards-based or competency-based grading. Most support the idea of reporting students’ achievement in terms of specific learning goals. They also understand the rationale behind giving students multiple opportunities to demonstrate what they have learned. The frustration of parents and families, as well as many students, comes from the failure of reform efforts to address what they consider a primary obstacle to fairness and equity in grading: *inconsistency in grading practices among teachers in the same school* (Guskey & Link, 2019). Each time students change classes, the rules for grading change. What counts as part of the grade, what doesn’t count, and how different aspects of students’ performance are weighed in determining grades — all can be different (Guskey, 2024).

This inconsistency leads many students to see grading as a game they must learn to play to succeed in school — and some students play the game quite well. They become strategists in the grading game, constantly tallying points and calculating the minimum scores they must attain to get the grade they want. But for other students, the grading game remains a mysterious puzzle they must decipher in every class, and many struggle in that effort. So, when a parent asks at the dinner table, “What grade are you going to get in this class?” the student responds in all honesty, “I don’t know.”

Before standards-based or competency-based grading reforms can be implemented, this inconsistency in grading must be addressed. This doesn't mean infringing on teachers' professional freedom. It simply requires reaching consensus about the purpose of grading and then implementing grading policies and practices that evidence shows serve the best interests of students and their learning.

Gaining greater consistency in grading among teachers involves three crucial steps that lay the groundwork for standards-based and competency-based grading reforms (Guskey, 2021):

1. Reach consensus on a clear and concise purpose statement for grading.
2. Use grading scales with four to seven categories of student performance.
3. Report academic and non-academic aspects of students' performance separately.

### **Develop a clear and concise purpose statement**

Teachers generally don't agree on why they give grades in the first place (Russell & Airasian, 2011). When neither teachers nor school leaders agree on what grades mean or what they are for, grading procedures tend to vary from teacher to teacher, class to class, and school to school.

#### ***Establishing consensus***

Successful grading reforms *always* begin with focused discussions on the purpose of grades and report cards (Brookhart, 2011). These discussions must address three questions:

1. What information will grades communicate?
2. Who is the primary audience for that information?
3. What is the intended goal of grading?

Reaching consensus on answers to these questions provides the foundation for determining the appropriateness of all grading policies and practices. It also establishes criteria for deciding the optimal form and structure of the report card.

Research by Jessica Gogerty (2016) showed that when the purpose of grading is clearly articulated, teachers become more deliberate in their approach to student learning. They prioritize curriculum standards and adjust their instructional procedures so that content, format, and difficulty of classroom assessments are more closely aligned. Teachers also express less tolerance of colleagues who fail to align their teaching and learning practices to the grading purpose. They see this failure as "negligence" that causes unnecessary confusion for students and families (p. 154). When the grading purpose is clear, teachers are expected to uphold that purpose.

### *Example purpose statements*

To guarantee a shared understanding among all stakeholders, this purpose statement should be prominently featured on the report card and included in the introduction of all grading policy documents. This helps clarify the report card's intent, the information it includes, and how to interpret that information.

Numerous examples of purpose statements for grading and report cards can be found online and in *Developing Standards-Based Report Cards* (Guskey & Bailey, 2010). Although these examples vary widely, the best succinctly address the three questions described earlier. An elementary-level example would be:

The purpose of this report card is to describe students' learning progress to parents and families, based on our school's learning goals for each grade level. It is intended to inform parents and families about learning successes and to guide improvements when needed.

This statement specifies the aim of the report card, for whom it is intended, and how the included information should be used. It is brief but clear and concise. Another example for the middle school or high school level is:

The purpose of this report card is to communicate with parents, families, and students about the achievement of specific learning goals. It identifies students' current levels of performance regarding those goals, areas of strength, and areas where additional time and effort are needed.

This statement identifies parents, families, *and* students as important audiences for the report card. It further specifies that the information describes students' "current level of performance," not where they started or an average of scores over time. It also indicates how the information should be used to guide improvement.

A third example comes from the American School of Paris, an international school where the administrators and faculty have been especially thoughtful in their approach to grading and reporting reform:

The primary purpose of grading is to effectively communicate student achievement toward specific standards, at this point in time. A grade should reflect what a student knows and is able to do. Students will receive separate feedback and evaluation on their learning habits, which will not be included in the academic achievement grades.

Two parts of this purpose statement deserve attention. First, the phrase "at this point in time" makes clear that teachers do not determine students' grades by averaging scores from the entire grading period. Instead, they assign grades based on the most current evidence they have on what students now know and can do. In other words, grades reflect where students are in their learning *right now*, not where they were weeks or months before.

Second, the statement "students will receive separate feedback and evaluation on their learning habits" emphasizes that achievement grades represent students' performance on specific

academic learning goals. Other aspect of students' behavior related to learning habits, such as homework completion, class participation, and punctuality in turning in assignments, are reported separately.

### **Use grading scales with four to seven performance categories**

Consistency in grading implies that teachers with comparable knowledge and experience, when presented with the same body of evidence on a student's performance, agree on the grade. Researchers call this "inter-rater reliability" (Gwet, 2021; Hallgren, 2012). The number of levels or categories of performance in the grading scale plays a significant role in achieving that agreement. Scales that include large numbers of categories increase the potential influence of subjectivity and drastically reduce agreement among teachers.

#### ***Direct and indirect measures***

The challenge of gaining acceptable levels of inter-rater reliability in grading is further complicated by the fact it requires teachers to summarize quantitative evidence gathered primarily through *indirect* approaches to measurement. *Direct* measurement involves explicitly measuring and quantifying the characteristic of a person that we want to report. For instance, to measure a student's height, we would ask the student to stand with their back against a wall, place a level instrument like a ruler or book on the top of their head, mark the wall, and then measure the distance from the floor to that mark. The recorded number represents the direct measurement of the student's height.

Most of the measures teachers use to determine grades are *indirect* measures. Indirect measurement involves measuring something else and converting it into a measurement of the characteristic in question (American Psychological Association, 2018). For example, we cannot directly measure students' achievement or proficiency by placing a measuring device on them. Instead, we ask students to answer questions or perform certain tasks. We then make *judgments* or *inferences* about students' level of achievement based on their responses or performance. Because these judgments involve personal interpretation, indirect measures are more susceptible to bias and interpretation errors than direct measures. This makes it extremely difficult to accurately discern and report subtle differences in students' performance.

When neither teachers nor school leaders agree on what grades mean or what they are for, grading procedures tend to vary from teacher to teacher, class to class, and school to school.

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

### ***The problem of the percentage scale***

In this context, the percentage grade scale with 101 discrete levels of student performance, two-thirds of which typically designate failure, presents a noteworthy challenge. Some educators believe the large number of levels in the percentage grade scale makes it more precise than scales with fewer levels, such as the five-level letter-grade scale (A, B, C, D, and F) used in most colleges and universities. But the reality is far more complex. In the absence of a truly accurate measuring device, adding more levels to the measurement scale offers only the illusion of precision. In fact, the large number of levels in the percentage scale, coupled with the fine discrimination required to determine differences among those levels using indirect measures, can lead to greater subjectivity, increased error, and diminished reliability. Researchers have recognized these problems for well over a century (Starch & Elliot, 1912, 1913).

In defense of the percentage grade scale, some educators argue that the percentage of questions on an assessment that students answer correctly represents a direct measure of achievement. They reason that correctly answering 80% of the questions on an assessment means the student has learned 80% of the material or mastered 80% of the learning goals. While the percentage of questions answered correctly might seem like a direct measure, the interpretation of this percentage involves numerous complexities. The format, difficulty, and alignment of the questions to instruction, as well as other factors, can significantly impact the accuracy and reliability of percentage-based scores. This complexity underscores the challenges in achieving true precision in grading, even when using seemingly straightforward measures like percentage correct. The perceived precision of percentage grading methods is far more illusory than real, due to the inherent subjectivity and complexity of the indirect measures involved (Guskey, 2013).

### ***Fewer levels, greater accuracy***

Significant research shows that optimal discrimination, validity, and reliability are obtained using grading scales with four to seven levels or categories (Lozano, Garcia-Cuento, & Muniz, 2008; Preston & Colman, 2000). Teachers with comparable knowledge and experience are far more likely to agree when distinguishing an A level from a B level of performance than when distinguishing a 90 from an 89 using the percentage scale. The use of clear and well-defined scoring criteria, along with a limited number of grading categories, helps ensure a shared understanding among teachers and promotes more consistent grading practices. This understanding is particularly important for implementing grading reforms that prioritize fairness, transparency, and equity.

## **Report multiple grades**

Every marking period, teachers gather multiple forms of evidence on students' performance that reflect three different types of grading criteria: *product*, *progress*, and *process* (Guskey, 1994, 1996).

- *Product* criteria show how well students have achieved specific academic learning goals, standards, or competencies, typically demonstrated through major assessments, classroom quizzes, compositions, projects, reports, and other culminating activities.
- *Progress* criteria, sometimes called “growth” or “development” criteria, show how much students have gained or improved in their learning. Students could make outstanding progress, but still not be achieving at grade level, and highly skilled students might achieve the product criteria without making notable improvement.
- *Process* criteria describe student behaviors that facilitate, broaden, or extend learning. These may include activities that enable learning, such as formative assessments, homework, and class participation. They also may reflect nonacademic social-emotional learning skills, such as collaboration, goal setting, perseverance, habits of mind, or citizenship. In some cases, they relate to students’ compliance with procedures, like turning in assignments on time.

### ***A hodgepodge grade***

At the end of each marking period, teachers assign weights to these different sources of evidence to tally a final score recorded on the report card (Sun & Cheng, 2013). Researchers call this a “hodgepodge” grade (Brookhart, 1991) because it mixes achievement and other factors related to behavior, attitude, effort, and improvement. It makes the report card grade a confusing amalgamation that is impossible to interpret clearly and accurately (Guskey, 2020). An A, for example, might mean that the student knew all the concepts before instruction began (product); that she did not achieve the learning goals but made significant improvement (progress); or that she put forth extraordinary effort (process).

Recognizing these problems, some grading reform advocates recommend that teachers use only product criteria in determining students’ grades. They point out that the more progress and process criteria come into play, the more subjective, biased, and inequitable grades become (Feldman, 2023). How can a teacher know, for example, how difficult a task was for students or how hard they worked to complete it? Many teachers point out, however, that if process elements like homework and punctuality in turning in assignments don’t count, students will lose all motivation to do homework or complete assignments on time — and evidence from schools implementing these practices confirm their apprehensions (Randazzo, 2023; We Are Teachers, 2023).

### ***Multiple grades for multiple criteria***

A far more effective solution is not to eliminate progress or process criteria from grading but to report these criteria separately. Teachers simply extract evidence on the important nonacademic aspects of students’ performance and report those in their own section of the report card *and* the transcript.

Although reporting multiple grades is relatively new in most U.S. schools, the practice has a long-established history in other countries. In Ontario, Canada, for example, teachers have reported multiple grades for students from 1st grade through high school for decades. Every

marking period, in addition to academic grades, teachers record grades for responsibility, independent work, initiative, organization, collaboration, and self-regulation. A major component of students' responsibility grade is "Completes and submits class work, homework, and assignments according to agreed-upon timelines." Students' grades for responsibility and other process elements are reported on a four-level scale with the categories *Excellent*, *Good*, *Satisfactory*, and *Needs Improvement* (Ontario Ministry of Education, 2023).

### ***Benefits for students, parents, teachers, and more***

Teachers using multiple grades say that knowing these aspects of performance will be reported on both the report card and transcript compels students to act more responsibly. Parents benefit because the report card provides a more detailed, comprehensive picture of their child's performance. In addition, because product grades are no longer tainted by evidence based on behavior or compliance, those grades more closely align with external measures of achievement and content mastery, such as standardized test scores — a quality college and university admissions officers favor (Buckmiller & Peters, 2018). In essence, removing process elements from the achievement (product) grade makes grades more accurate, honest, and equitable indicators of student learning.

Most important, reporting multiple grades doesn't require extra work for teachers. In fact, it's less work. Teachers already gather evidence on product, progress, and process criteria. For example, most keep records of students' scores on various measures of achievement, as well as homework completion, class participation, collaboration in projects, and so on. By simply reporting separate grades for these different aspects of learning, teachers avoid the dilemmas involved in determining how much to weigh each element when calculating a single grade.

### **A more accurate picture**

Establishing greater consistency in grading policies and practices doesn't require all teachers to grade in the same way. Just as assessment strategies must be adapted to fit the learning goals in different subjects, grading procedures must be similarly adapted to accurately communicate students' achievement of those learning goals.

Schools where educators reach consensus on a purpose statement, adopt a grading scale with four to seven categories of student performance, and report academic achievement and nonacademic learning goals separately have the necessary foundation for more meaningful and effective grading reform. With these three crucial steps accomplished, most teachers find it easy to transition to standards-based or competency-based grading. They recognize how they can break down an overall achievement grade to report on the different standards that it summarizes. Many see this transition as a natural progression in their efforts to provide meaningful summaries of students' performance. Without adding to teachers' workload, these steps address the greatest concerns of parents and families; facilitate better communication between school and home; and ensure greater honesty, accuracy, and equity in grading.



## References

- American Psychological Association. (2018). Indirect measurement. In *APA dictionary of psychology*. <https://dictionary.apa.org/indirect-measurement>
- Brookhart, S.M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10 (1), 35-36.
- Brookhart, S.M. (2011). Starting the conversation about grading. *Educational Leadership*, 69 (3), 10-14.
- Buckmiller, T.M. & Peters, R.E. (2018). Getting a fair shot? *School Administrator*, 75 (2), 22-25.
- Feldman, J. (2023). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms* (2nd ed.). Corwin.
- Franklin, A., Buckmiller, T., & Kruse, J. (2016). Vocal and vehement: Understanding parents' aversion to standards-based grading. *International Journal of Social Science Studies*, 4 (11), 19-29.
- Gogerty, J.I. (2016). *The influence of district support during implementation of high school standards-based grading practices* [Unpublished doctoral dissertation]. Drake University, Des Moines, Iowa.
- Guskey, T.R. (1994). Making the grade: What benefits students. *Educational Leadership*, 52 (2), 14-20.
- Guskey, T.R. (1996). Reporting on student learning: Lessons from the past — Prescriptions for the future. In T.R. Guskey (Ed.), *Communicating student learning* (pp. 13-24). ASCD.
- Guskey, T.R. (2013). The case against percentage grades. *Educational Leadership*, 71 (1), 68-72.
- Guskey, T.R. (2020). Breaking up the grade. *Educational Leadership*, 78 (1) 41-46.
- Guskey, T.R. (2021). Learning from failures: Lessons from unsuccessful grading reform initiatives. *NASSP Bulletin*, 105 (3), 192-199.
- Guskey, T.R. (2024). *Engaging parents and families in grading reforms*. Corwin.
- Guskey, T.R. & Bailey, J.M. (2010). *Developing standards-based report cards*. Corwin.
- Guskey, T.R. & Brookhart, S.M. (2019). *What we know about grading: What works, what doesn't, and what's next*. ASCD.
- Guskey, T.R. & Link, L.J. (2019, April). *Understanding different stakeholders' views on homework and grading* [Paper presentation]. Annual Meeting of the American Educational Research Association, Toronto, ON, Canada.

- Gwet, K.L. (2021). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (5th ed.). AgreeStat Analytics.
- Hallgren K.A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychol*, 8 (1), 23-34.
- Jansen, B.J., Salminen, J., Jung, S., & Almerexhi, H. (2022). The illusion of data validity: Why numbers about people are likely wrong. *Data and Information Management*, 6 (4), 1-14.
- Lozano, L.M., Garcia-Cuento, E., & Muniz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4 (2), 73-79.
- Ontario Ministry of Education. (2023). *Elementary and secondary report card templates*. [www.ontario.ca/page/elementary-and-secondary-report-card-templates](http://www.ontario.ca/page/elementary-and-secondary-report-card-templates)
- Preston, C.C. & Colman, A.M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104 (1), 1-15.
- Randazzo, S. (2023, April 26). Schools are ditching homework, deadlines in favor of “Equitable grading.” *Wall Street Journal*.
- Russell, M.K. & Airasian, P.W. (2011). Grading. In *Classroom assessment: Concepts and applications* (7th ed.). McGraw-Hill.
- Starch D. & Elliott, E.C. (1912). Reliability of the grading of high school work in English. *School Review*, 20, 442-457.
- Starch, D. & Elliott, E.C. (1913). Reliability of the grading of high school work in mathematics. *School Review*, 21, 254-259.
- Sun, Y. & Cheng, L. (2013). Teachers’ grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21, 326-343.
- We Are Teachers Staff. (2023, September 7). “No zeros” is sold as an equity shortcut. It’s not. We Are Teachers. [www.weareteachers.com/equitable-grading/](http://www.weareteachers.com/equitable-grading/).
- Young, J. (2023, November 30). Some schools are changing how they grade students. Here’s why some parents are upset. *USA Today*.
- This article appears in the May 2024 issue of *Kappan*, Vol. 105, No. 8, p. 52-57.

## **ABOUT THE AUTHOR**

**Thomas R. Guskey** is professor emeritus at the University of Kentucky, Lexington. He is the author of *Get Set, Go! Creating a Successful Grading and Reporting System*.