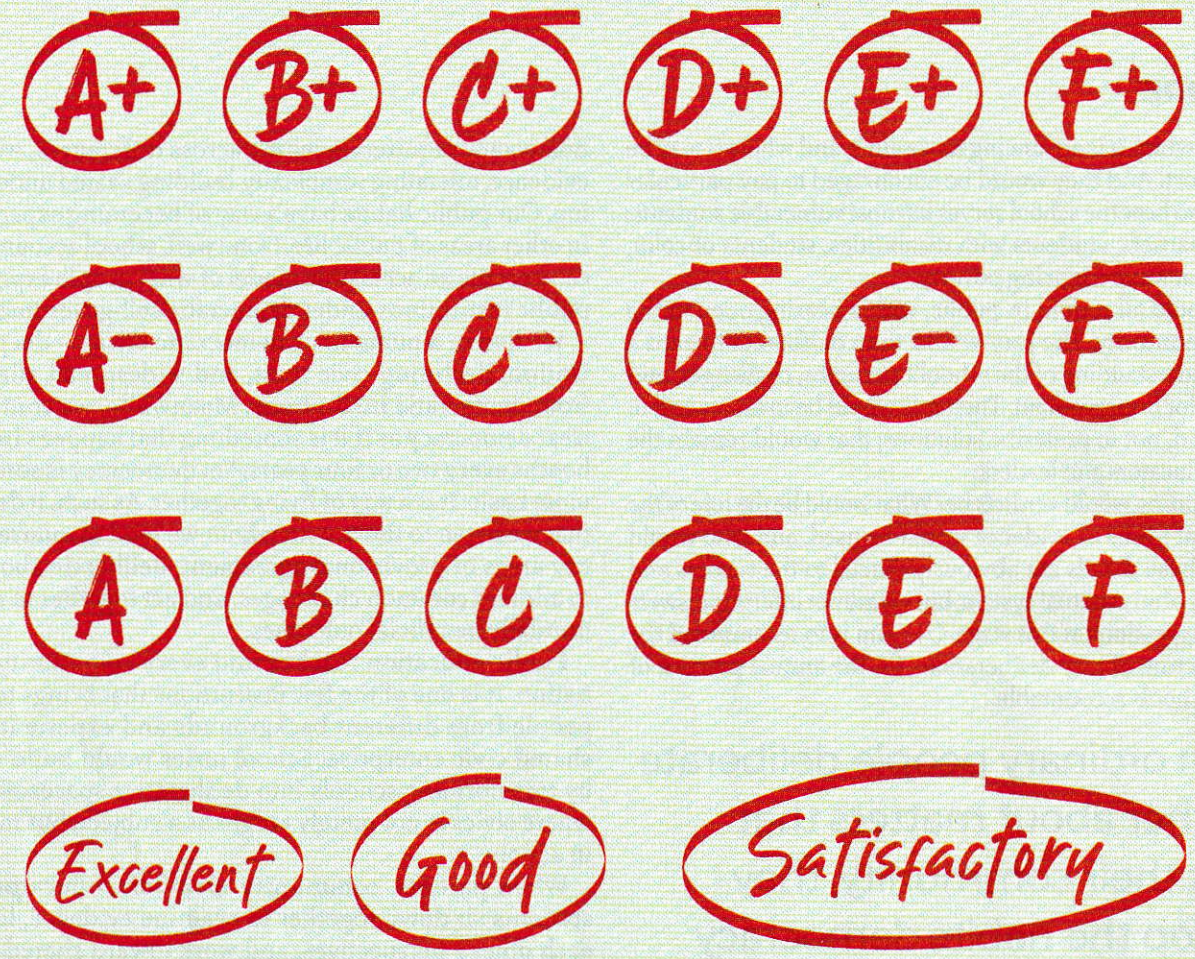


TOWARD CLARITY AND COHERENCE IN STANDARDS-BASED GRADING

A lack of consistency, terminology, and messaging has added to the difficulty of reforming a flawed grading system. A clear foundation to build on is needed.

By Thomas R. Guskey

Over the last decade, interest in grading and reporting student learning has exploded. Of the 129 books released on this topic over the past six decades, more than two-thirds were published in just the last 15 years (Guskey, Frey, and Fisher, 2024). This surge has been accompanied by a significant increase in research on the topic (Guskey & Brookhart, 2019).

Although the authors of these recent books offer well-intentioned proposals for reforming grading policies and practices, they share little agreement on what changes should be made or where reform efforts should begin. Key terms like "standards-based," "competency-based," and "standards-referenced" grading are defined in markedly different ways. In addition, few authors offer

GETTY IMAGES

THOMAS R. GUSKEY (guskey@uky.edu) is professor emeritus in the College of Education at the University of Kentucky, Lexington.

reliable evidence to confirm that their recommendations enhance communication between schools and families or improve student outcomes (Link & Guskey, 2022). The only point on which most seem to agree is that traditional grading practices are deeply flawed.

This lack of coherence leaves school leaders uncertain about who to trust and what to believe. Confronted with conflicting messages, most struggle to implement grading reforms, often encountering unexpected controversy and resistance. In many cases, leaders find themselves managing problems their reform efforts inadvertently created (Larson et al., 2024) and frequently are forced to substantially revise their plans or abandon reform efforts altogether.

To implement effective, sustainable grading reforms that benefit students, parents, teachers, and school leaders alike, we must establish shared terminology and recognize the crucial prerequisites to successful reform initiatives

What is standards-based grading?

Standards-based grading (SBG) is a system in which students' achievement and learning progress are evaluated and reported according to how well they have met clearly articulated learning goals (Tomlinson & McTighe, 2006). Although some authors differentiate between "standards" and "competencies" (Hess, Colby, & Joseph, 2020), most SBG advocates follow the guidance of Ralph Tyler (1949), who maintained that while educators must be explicit about what students are expected to learn *and* be able to do, the specific label attached to those learning goals — be it standards, competencies, outcomes, proficiencies, targets, or objectives — is largely inconsequential.

The defining features of SBG were first articulated by Megan Welsh and Gerald D'Agostino (see D'Agostino & Welsh, 2007; Welsh & D'Agostino, 2009; Welsh & Eastwood, 2011; Welsh, D'Agostino, & Kaniskan, 2013) and later refined by Welsh (2019). These features are:

1. Student achievement is reported in relation to key grade-level or course standards rather than as a single, aggregated subject-area grade.
2. Student achievement is communicated using a limited set of performance categories, typically four to seven, that describe students' progress toward mastery of specific learning goals.
3. Student academic achievement grades are reported separately from nonacademic or behavioral factors such as effort, homework completion, and class participation.

Retakes and SBG

Some proponents of SBG add a fourth feature of allowing students multiple opportunities to demonstrate their proficiency (see Beatty, 2013; Townsley & Buckmiller, 2016; Townsley & Wilcox, 2024). However, this practice relates more to assessment policies than it does to grading and reporting. It derives from the instructional practices

AT A GLANCE

- Authors and researchers define standards-based grading (SBG) and associated terms in a variety of ways, creating misunderstandings and making grading reform difficult.
- According to research, SBG can be defined as a system in which students' achievement and learning progress are evaluated and reported according to how well they have met clearly articulated learning goals.
- Retakes can be part of SBG systems, but they are best implemented as part of a mastery learning system.
- Transitioning to SBG requires schools to establish a purpose for grading, use grading scales with four to seven categories, and report academic and non-academic aspects of students' performance separately.

associated with mastery learning in which teachers use regular formative assessments to offer students feedback and guide the correction of learning difficulties (see Bloom, 1968; Guskey, 2023a).

Problems inevitably arise when retakes are incorporated into grading without the instructional components emphasized in mastery learning (Guskey, 2023b). For instance, when students are given the chance to retake assessments without participating in the teacher-led corrective or enrichment activities that are central to mastery learning, teachers often have difficulty motivating students to put forth their best effort on the initial attempt (Coffey & Tyner, 2024; Nijenkamp et al., 2022).

Some school districts attempt to address this problem by placing a cap on the maximum score students can earn on a retake. In other words, they may stipulate that students who require a retake can receive no grade higher than a B (see Fairfax Public Schools, 2025). But this approach sends a troubling message to students. Think of it this way:

> The learning goal is to clear a one-meter hurdle. Students who do not succeed on their first attempt are allowed to practice and try again. Yet even if they clear the one-meter hurdle on their second attempt, they will be credited only for clearing a 90-centimeter hurdle.

Such a grading policy is not only inequitable; it also misrepresents students' actual learning and undermines the fundamental purpose of grading.

Crucial prerequisites to grading reform

When parents and families are asked about their greatest concern regarding fairness in grading, they most frequently identify inconsistencies in grading practices among teachers

within the same school (Guskey & Link, 2019). As students move from one class to another, they often encounter different grading rules. What counts toward a grade, what does not, and how various aspects of performance are weighted can vary widely from teacher to teacher (Guskey, 2024a). These inconsistencies cause many students to perceive grading as a game they must learn to play to succeed. As a result, students become strategists, focusing on point accumulation and calculating the minimum scores needed to achieve their desired grade (Guskey, 2024b).

For standards-based grading reforms to be implemented effectively, this lack of consistency must be addressed. Resolving it does not limit teachers' professional autonomy. Rather, it requires meeting three essential conditions that promote consistent grading policies and practices within a school (Guskey, 2021):

1. Establish a clear and concise purpose for grading.
2. Use grading scales with four to seven performance categories.
3. Report academic achievement and nonacademic factors separately.

Two of these conditions align directly with Welsh's (2019) defining criteria of standards-based grading.

Establish a clear and concise purpose for grading

Successful grading reform begins with intentional conversations about the purpose of grades and report cards (Brookhart, 2011). These discussions focus on three essential questions:

1. What information should grades communicate?
2. Who is the intended audience?
3. What is the ultimate goal of reporting student performance?

Reaching agreement on these questions establishes the foundation for evaluating grading policies and practices. It also guides decisions about the most effective structure and format of the report card. When the purpose of grading is clearly defined, teachers focus more deliberately on curriculum standards and align instructional practices and classroom assessments in terms of content, format, and difficulty (Gogerty, 2016).

An excellent example of a purpose statement was developed by the faculty and school leaders at the American School of Paris, an international school that has implemented successful grading reforms:

> 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 have special significance.

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 about what students currently know and can do. In other words, grades reflect where students are now in their learning, not where they were weeks or months before.

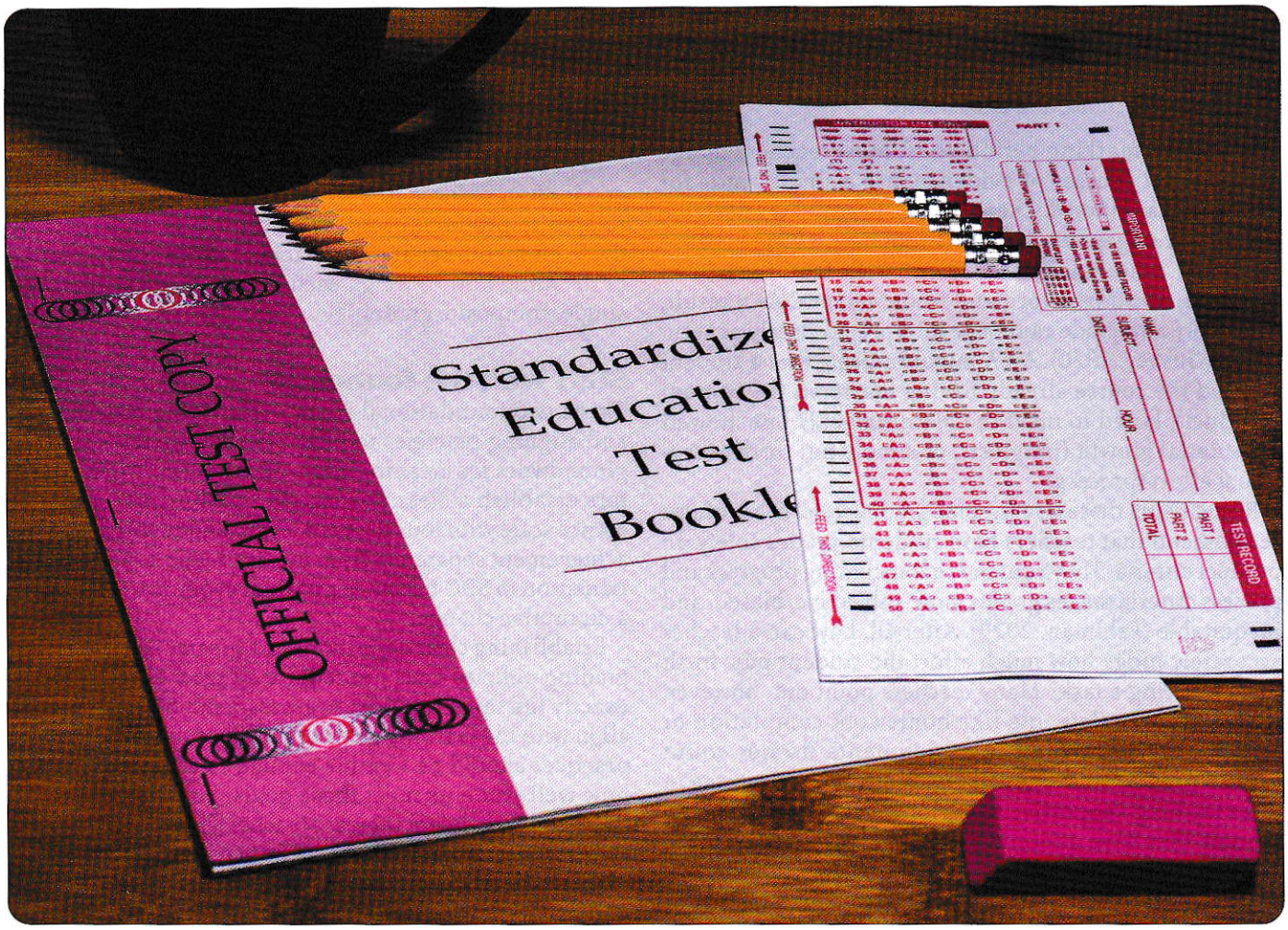
Second, the statement "students will receive separate feedback and evaluation on their learning habits" emphasizes that achievement grades represent students' performance on specific academic learning goals. Other aspects of students' behavior related to learning habits, such as homework completion, class participation, and punctuality in turning in assignments, are reported separately.

Standards-based grading (SBG) is a system in which students' achievement and learning progress are evaluated and reported according to how well they have met clearly articulated learning goals

Use grading scales with four to seven performance categories

Consistency in grading requires that when teachers with comparable knowledge and experience are presented with the same evidence of a student's performance, they agree on the grade. Researchers describe this agreement as *inter-rater reliability* (Gwet, 2021; Hallgren, 2012). One important factor influencing inter-rater reliability is the number of performance levels included in the grading scale. In this context, the percentage grading scale, which contains 101 discrete levels of performance, with nearly two-thirds typically representing failure, poses a significant challenge.

Some educators argue that the large number of levels in the percentage scale makes it more precise than systems with fewer categories, such as the five-level letter-grade scale (A, B, C, D, and F). In practice, however, greater complexity does not necessarily lead to greater accuracy. In the absence of a truly precise measuring instrument, adding more levels to a scale creates only the illusion of precision. In fact, the fine distinctions required by the percentage scale often increase subjectivity, reduce reliability, and introduce greater measurement error. Researchers have documented these problems for more than a century (Starch & Elliot, 1912, 1913).



In defense of the percentage grade scale, some educators maintain that the proportion of questions answered correctly on an assessment provides a direct measure of student achievement. In short, they assume that answering 80% of the questions correctly indicates mastery of 80% of the content. While this interpretation may seem straightforward, it overlooks important complexities. The format, difficulty, and instructional alignment of assessment questions, as well as other contextual factors, can significantly influence the accuracy and reliability of percentage-based scores (Guskey, 2013).

Extensive research shows that grading scales with four to seven performance levels yield optimal discrimination, validity, and reliability (Lozano, Garcia-Cuento, & Muniz, 2008; Preston & Colman, 2000). Teachers are far more likely to agree when distinguishing between an A and a B level of performance than when attempting to differentiate between a score of 90 and 89 on a percentage scale. Using a limited number of clearly defined grading categories, supported by explicit scoring criteria, promotes shared understanding among educators and leads to more consistent and reliable grading practices.

Report academic and nonacademic aspects of students' performance separately

During each marking period, teachers collect multiple forms of evidence of student performance that reflect three distinct types of grading criteria: *product, progress,* and *process* (Guskey, 1994, 1996).

- *Product* criteria indicate how well students have achieved specific academic learning goals and are typically demonstrated through major assessments such as tests, quizzes, writing assignments, projects, reports, and other culminating tasks.
- *Progress* criteria, sometimes referred to as growth or development, show the extent to which students have improved in their learning over time. A student may make significant progress yet still perform below grade level, while another may meet achievement expectations with little growth.
- *Process* criteria describe behaviors and dispositions that support learning, including formative

GETTY IMAGES

assessments, homework completion, and class participation, as well as social-emotional competencies such as collaboration, perseverance, goal setting, habits of mind, and citizenship.

When the marking period ends, most teachers merge these varied sources of evidence into a single grade for the report card (Sun & Cheng, 2013). Researchers describe this as a "hodgepodge" grade (Brookhart, 1991) because it blends academic achievement with factors such as behavior, attitude, effort, and improvement. As a result, the report card grade cannot be interpreted clearly or accurately (Guskey, 2020). An A might indicate that a student entered the course already having mastered the content (product), failed to meet the learning goals but showed substantial growth (progress), or simply put forth exceptional effort (process).

Recognizing these issues, some grading reform advocates recommend that teachers base students' grades solely on product criteria. They contend that including progress and process criteria makes grades more subjective, biased, and inequitable (Feldman, 2023). After all, how can a teacher accurately judge how much effort the student puts forth in completing a task? Many teachers point out, however, that if process elements like homework completion or punctuality in turning in assignments do not count toward grades, students will lose motivation to engage in these learning-supportive behaviors — and evidence from schools that have adopted such policies supports these concerns (Griffith & Tryner, 2025; Randazzo, 2023).

A more effective approach is not to remove progress or process criteria from grading, but to report them separately. Teachers can collect evidence on key nonacademic aspects of student performance and report those in their own section of the report card and transcript.

While relatively new in most U.S. schools, this 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 using a four-level scale with the categories *Excellent, Good, Satisfactory,* and *Needs Improvement* (Ontario Ministry of Education, 2023). Reporting multiple grades also aligns with programs currently underway in many states to develop "portraits of a graduate" (Atwell & Tucker, 2024).

Teachers who use multiple grades report that when students know these different aspects of performance will appear on both the report card and transcript, they tend to act more responsibly. Parents also benefit because the report card offers a more detailed, comprehensive picture of their child's performance. In addition, separating behavior- and compliance-related evidence from product grades results in achievement grades that more accurately reflect content mastery and align better with external measures such as standardized test scores — an outcome valued by college and university admissions officers (Buckmiller & Peters, 2018; Welsh & D'Agostino, 2009).

Reporting multiple grades does not add work for teachers; it often reduces it. Teachers already collect evidence related to product, progress, and process. By reporting these elements separately, teachers avoid the challenge of deciding how to weigh each factor when calculating a single, composite grade.

Start with the foundation

These two foundational steps — clarifying terminology and aligning practices to promote consistency — lay the groundwork for meaningful grading reform. When educators establish a clear purpose for grading, use a performance scale with four to seven levels, and report academic achievement separately from nonacademic behaviors, the transition to SBG becomes a natural evolution rather than a disruptive change.

Establishing clear definitions and greater consistency in grading policies does not require all teachers to grade in exactly the same way. Just as assessment methods vary to align with the learning goals of different subjects, grading practices should be flexible enough to accurately reflect how well students meet those goals. Equally important, these changes do not increase teachers' workload. Instead, they address common concerns from parents and families, strengthen communication between home and school, and promote greater honesty, accuracy, and fairness in grading practices.

Once this foundation is in place, most teachers find the transition to SBG relatively straightforward. They quickly see how a single overall grade can be disaggregated into specific standards, resulting in more precise and informative communication. For many educators, this transition feels like a logical extension of their ongoing efforts to provide clearer, more actionable feedback. Furthermore, it addresses common concerns from parents and families about grading; strengthens communication between home and school; and promotes greater honesty, accuracy, and fairness in grading practices. ■

References

Atwell, M.N., & Tucker, A. (2024, February). *Portraits of a graduate: Strengthening career and college readiness through social and emotional skill development* (ERIC No. ED641286). Collaborative for Academic, Social, and Emotional Learning.

Beatty, I.D. (2013). Standards-based grading in introductory university physics. *Journal of the Scholarship of Teaching and Learning, 13* (2), 1-22.

Bloom, B.S. (1968). Learning for mastery. *Evaluation Comment* (UCLA-CSIEP), *1* (2), 1-12.

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.

Coffey, M. & Tyner, A. (2024, February). *Think again: Does "equitable" grading benefit students?* Thomas B. Fordham Institute.

D'Agostino, J.V. & Welsh, M.E. (2007, April). Standards-based progress reports and standards-based assessment score convergence. In T. R. Guskey (Chair), *The challenges of standards-based grading and reporting.* Symposium conducted at the meeting of the American Educational Research Association, Chicago, Illinois.

Fairfax County Public Schools (2025). *Middle and high school grading and reporting handbook.* https://www.fcps.edu/academics/grading-and-reporting/secondary-school.

Feldman, J. (2023). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms* (2nd ed.). Corwin Press.

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, IA.

Griffith, D. & Tryner, A. (2025). *"Equitable" grading through the eyes of teachers.* Thomas B. Fordham Institute.

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. (2023a). *Implementing mastery learning* (3rd ed.). Corwin Press.

Guskey, T.R. (2023b). Giving retakes their best chance to improve learning. *Educational Leadership, 80* (9).

Guskey, T.R. (2024a). *Engaging parents and families in grading reforms.* Corwin Press.

Guskey, T.R. (2024b). Addressing inconsistencies in grading practices. *Phi Delta Kappan, 105* (8), 52-57.

Guskey, T.R. & Brookhart, S.M. (Eds.) (2019). *What we know about grading: What works, what doesn't, and what's next?* ASCD.

Guskey, T.R., Frey, N., & Fisher, D. (2024). *Grading with integrity.* Corwin Press.

Guskey, T.R. & Link, L.J. (2019, April). *Understanding different stakeholders' views on homework and grading.* Paper presented at the 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. *The Quantitative Methods for Psychology, 8* (1), 23-34.

Hess, K., Colby, R., & Joseph, C. (2020). *Deeper competency-based learning.* Corwin Press.

Larson, G., Peppler, B., Smith, D., & Townsley, M. (2024). *Extinguishing the fires within assessment and grading reform.* FIRST Education Resources.

Link, L.J., & Guskey, T.R. (2022). Is standards-based grading effective? *Theory Into Practice, 61* (4), 406-417.

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.

Nijenkamp, R., Nieuwenstein, M.R., de Jong, R., Lorist, M.M. (2022). Second chances in learning: Does a resit prospect lower study-time investments on a first test? *Journal of Cognition, 5* (1), 1-23.

Ontario Ministry of Education (2023). *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.' *The Wall Street Journal.*

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.

Tomlinson, C.A. & McTighe, J. (2006). *Integrating differentiated instruction & understanding by design: Connecting content and kids.* ASCD.

Townsley, M. & Buckmiller, T. (2016, January). *What does the research say about standards-based grading?* http://mctownsley.net/standards-based-grading-research

Townsley, M. & Wilcox, J. (2024). Standards-based grading: Core principles and current implementation debates. *The Clearing House: A Journal of Educational Strategies, Issues, and Ideas, 97,* 76-86.

Tyler, R.W. (1949). *Basic principles of curriculum and instruction.* University of Chicago Press.

Welsh, M.E. (2019). Standards-based grading. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading: What works, what doesn't, and what's next* (pp. 133-144). ASCD.

Welsh, M.E. & D'Agostino, J. V. (2009). Fostering consistency between standards-based grades and large-scale assessment results. In T.R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 75-104). Corwin Press.

Welsh, M.E., D'Agostino, J.V., &. Kaniskan, R.B. (2013). Grading as a Reform Effort: Do Standards-Based Grades Converge with Test Scores? *Educational Measurement: Issues and Practice, 32* (2), 26-36.

Welsh, M.E. & Eastwood, M. (2011, April). *Conceptualizing teaching to the test under standards-based reform.* Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.