JSARD

ISSN: 2470-850X
Online

JOURNAL OF SCHOOL ADMINISTRATION RESEARCH & DEVELOPMENT

SPECIAL ISSUE: CURRENT AND FUTURE GRADING PRACTICES IN K-12 SCHOOLS

Vol. 11, S1
Summer 2026

ANNIVERSARY EDITION

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 51-53
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

Introduction to the Special Issue: Exploring Current and Future Classroom Grading Practices in K-12 Schools

Matt Townsley
Special Issue Guest Editor
University of Northern Iowa, USA

Since its inception in 2016, the *Journal of School Administration Research and Development (JSARD)* has focused on a range of topics within school leadership and administration, including grading and assessment. In this special issue, *JSARD* readers will benefit from understanding where the field of grading has been in the past ten years and where it may be headed in the future. Through invited and contributed manuscripts, *JSARD* continues to position itself as an established open-access journal empowering researchers, practitioners, and policymakers to improve K-12 schools.

BACKGROUND OF THIS SPECIAL ISSUE

Grading in K-12 schools is one of the most private experiences for teachers and has been called "the third rail" in education because it is so deeply rooted in tradition (Erickson, 2010). Brookhart and colleagues (2016) reviewed over a century of grading research and found that "grades assess a multidimensional construct containing both cognitive and non-cognitive factors reflecting what teachers value in student work" (p. 803). In response to these all-encompassing grading practices, Link and Guskey (2022) have reported that more schools are implementing standards-based grading to better communicate how well students have mastered academic standards. A search within the Institute of Education Sciences ERIC database suggests that the number of researchers investigating standards-based grading practices has sharply increased in the past 10 years.

Since 2016, when Brookhart and colleagues published their summary of grading research in *Review of Educational Research (RER)*, the world has experienced a global pandemic that induced emergency remote learning for schools across the globe. Stemming from these rapid changes to the student educational experience, state departments of education provided guidance to districts on their guiding principles, student advancement, and how to determine letter grades (Townsley & Kunnath, 2022). During the peak of COVID-19, some schools enacted "do no harm grading" practices that provided students with more flexibility in their submission deadlines and grace in the learning process (Townsley, 2020). A strong focus on equity within grading has emerged (Feldman, 2019; 2020), with a number of schools changing their grading practices in line with other more equitable educational practices and policies (Candelaria, 2022; King, 2022). While preliminary research suggests many of these emergency-influenced grading shifts may be temporary, flexible grading practices appear to remain relevant beyond the pandemic (Kunnath et al., 2026). Celebrating the tenth anniversary of *JSARD* and one decade following the publication of Brookhart and colleagues' *RER* paper, the current issue provides insights on the condition of K-12 grading practices in 2026 and beyond.

EXPLORING CURRENT AND FUTURE GRADING PRACTICES IN K-12 SCHOOLS

This special issue includes research articles, scholarly and critical essays, and commentaries that explore the challenges and opportunities in the field of K-12 grading over the past 10 years and in the years ahead. The articles in this special issue also offer practical knowledge and guidance for school leaders and provide researchers and policymakers with an enhanced theoretical understanding of grading practices in K-12 schools.

In the feature article of this special issue, renowned assessment and grading scholar Susan Brookhart presents an essay exploring additions to the grading research literature observed in the decade since the comprehensive review in 2016 (Brookhart et al., 2016). Researchers in the field will appreciate this update as an opportunity to consider the effects of the COVID-19 pandemic, a continued focus on standards-based grading within the literature, and future directions for grading research. I am personally grateful to Dr. Brookhart for contributing this timely essay describing a growing interest in grading practices, while also challenging researchers to continue exploring this important topic empirically.

Collectively, the articles in this special issue illuminate the promise and persistent challenges of grading reform. Link and colleagues (2026) add a broad stakeholder perspective to the conversation through survey data from more than 20,000 participants across 10 school districts, revealing substantial gaps among administrators, teachers, parents, support staff, and students regarding the role of behavior in grades. Their findings highlight the need for greater stakeholder alignment, student voice, and a continued focus on equity-centered grading policies. Examining grading practices from a different perspective, Kollman and colleagues (2026) analyze high school gradebook data and find frequent discrepancies between students' summative assessment performance and final course grades, with grade inflation occurring more often than grade deflation. Their results raise important questions about the accuracy and fairness of traditional point-based grading systems. Beyond K–12 classrooms, VanDevender (2026) investigates how undergraduate teacher education programs prepare future educators for standards-based grading, proposing a continuum that ranges from exposure to fully aligned programmatic integration. Together, these articles underscore the interconnected nature of grading reform, spanning policy, classroom practice, and educator preparation.

The special issue also explores implementation, innovation, and alternatives to traditional grading systems. Percell (2026) offers a cautionary case study of a district that abandoned standards-based grading after a decade-long effort, demonstrating how even carefully planned reforms can falter when implementation challenges, communication difficulties, and technological limitations undermine long-term support. Carney and Shume (2026) contribute a conceptual analysis of methods used to determine proficiency on individual standards, comparing literature-informed approaches and introducing Boolean Logic Gates as a novel framework for communicating student learning. Piantaggini (2026) examines the notion of grade "weaponization" through a practitioner-research lens, documenting how ungrading, self-assessment, flexible deadlines, and retake opportunities may reduce the harmful effects of permanent grades and better support student learning.

Finally, *JSARD* editor-in-chief Josh Kunnath and I (2026) propose a 5-P framework for school leaders to start and sustain grading reform. Our framework starts with establishing a guiding coalition that develops an understanding of why grading needs to change, followed by a multi-step process of establishing grading policies, implementing these grading policies over time, and providing sustained support through feedback loops. We believe school leaders will benefit from this framework as they guide educators in their local context with practical steps beyond attending a grading workshop or participating in a book study.

Collectively, the eight articles in this special issue advance the field's understanding of grading reform by addressing fundamental questions of equity, accuracy, implementation, and teacher preparation, while offering empirical evidence and practical pathways for reimagining how grades can better support student learning in 2026 and beyond. I hope you enjoy reading this special issue that explores current and future classroom grading practices in K-12 schools!

REFERENCES

Brookhart, S.M. (2026). Changes in K-12 grading research since 2016. *Journal of School Administration Research and Development, 11*(S1), 54-62. https://doi.org/10.32674/yvxznm80 *

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. https://doi.org/10.3102/0034654316672069

Candelaria, E. (2022, November). APS dips its toe into equitable grading. *Albuquerque Journal*. https://www.abqjournal.com/2552844/ex-pilot-program-focuses-on-the-students-mastery-of-a-subject.html

Carney, S., & Shume, T. (2026). Boolean logic gates: A new method of grade calculation for standards-based grading. *Journal of School Administration Research and Development, 11*(S1), 121-130. https://doi.org/10.32674/6j6j7f90 *

Erickson, J.A. (2010). Grading practices: The third rail. *Principal Leadership, 10*(7), 22-26.

Feldman, J. (2019). Beyond standards-based grading: Why equity must be part of grading reform. *Phi Delta Kappan, 100(*8), 52-55. https://doi.org/10.1177/0031721719846890

Feldman, J. (2020). Taking the stress out of grading. *Educational Leadership, 78*(1), 14-20.

King, G. (2022, November). Fewer D's and F's at Liberty High with equitable grading. *Cedar Rapids Gazette*. https://www.thegazette.com/k/fewer-ds-and-fs-at-liberty-high-with-equitable-grading/

Kollman, A., Sinner, E., Nurmi, H., & O'Shea, C. (2026). A matter of standards: Addressing inconsistencies in grading practices in high school English, mathematics, science, and social studies courses. *Journal of School Administration Research and Development, 11*(S1), 105-120. https://doi.org/10.32674/ftm8de57 *

Kunnath, J., Townsley, M., & Radunzel, J. (2026). Silver lining or missed opportunity? A mixed methods study of COVID-19 school closures and grading reform efforts in K-12 schools. *Journal of School Leadership.* Online First. https://doi.org/10.1177/10526846261428221

Link, L. J. & Guskey, T. R. (2022). Is standards-based grading effective? *Theory into Practice, 61*(4), 406-417. https://doi.org/10.1080/00405841.2022.2107338

Link, L. J., Simonds, R. M., & Navarrete-Burks, L. A. (2026). Disrupting the hidden costs of behavioral grading: Uncovering stakeholder perceptions to advance equity. *Journal of School Administration Research and Development, 11*(S1), 74-83. https://doi.org/10.32674/1db46k56 *

Percell, J. (2026). The retrograde pendulum: Reversing course on standards-based grading. *Journal of School Administration Research and Development, 11*(S1), 84-89. https://doi.org/10.32674/7an6dt57 *

Piantaggini, L. (2026). De-weaponizing grades through teacher tinkering and student insight. *Journal of School Administration Research and Development, 11*(S1), 131-141. https://doi.org/10.32674/bs6x1g46 *

Townsley, M., & Kunnath, J. (2022). Exploring state department of education grading guidance during COVID-19: A model for future emergency remote learning. *Education Policy Analysis Archives, 30*,(163). https://doi.org/10.14507/epaa.30.7448

Townsley, M., & Kunnath, J. (2026). The 5 P's of starting and sustaining grading reform. *Journal of School Administration Research and Development, 11*(S1), 63-73. https://doi.org/10.32674/q6wpb124 *

Townsley, M. (2020). Grading principles in pandemic-era learning: Recommendations and implications for secondary school leaders. *Journal of School Administration Research and Development, 5*(S1), 8-14. https://doi.org/10.32674/jsard.v5iS1.2760

VanDevender, K. J. (2026). From fragmented to aligned: A standards-based grading continuum in teacher preparation. *Journal of School Administration Research and Development, 11*(S1), 90-104. https://doi.org/10.32674/zar92s63 *

**Denotes an article published in this special issue*

MATT TOWNSLEY, EdD, is an Associate Professor of Educational Leadership in the Department of Learning, Leadership and Community at the University of Northern Iowa. His current research interests include educational leadership and classroom grading practices. Email: matt.townsley@uni.edu

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 54-62
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

Changes in K-12 Grading Research Since 2016

Susan M. Brookhart
Duquesne University, USA

ABSTRACT

This essay explores additions to the grading research literature observed in the decade since the large review "A Century of Grading Research: Meaning and Value in the Most Common Educational Measure" (Brookhart et al., 2016). Sections discuss studies of traditional grading, the effects of the pandemic on grading, studies of standards-based grading, and studies about rubrics. A final section discusses future directions for grading research.

***Keywords*:** grading, standards-based grading, classroom assessment

In 2016, a group of colleagues and I published a review of the grading research literature (Brookhart et al., 2016), which at that point included over a century of study. For this special issue of JSARD on current and future grading practices in K-12 schools, I have been asked to comment on what I have observed in the field since that paper was published and where the field might be going in the future. I make no claim that this paper presents a systematic review of all the grading research in the last decade, which is beyond the scope of this commentary. However, I am happy to take up the two-fold challenge that was put to me and present thoughts on grading research since 2016 and possible future directions.

GRADING RESEARCH SINCE 2016

The 2016 paper included five sections: (a) early studies of the reliability of grades, (b) quantitative studies of the composition of K–12 report card grades, (c) survey and interview studies of teachers' perceptions of grades, (d) studies of standards-based grading (SBG), and (e) grading in higher education. Leaving aside the early historical studies and the higher education studies (because of the K-12 focus of this issue), we can identify K-12 empirical studies in the other three categories published since 2016: studies of traditional grading (b and c above), the effects of the pandemic on grading (c and d), and studies of SBG (d).

Recent Research on Traditional Grading

Recent (post-2016) research on traditional grading seems to me to extend what has gone before, either through increased methodological sophistication or through increased theoretical grounding. Two examples of sophisticated statistical modeling include Doz (2023) and Klapp (2017).

Doz (2023) used a large national dataset from Italy and found that teacher-assigned mathematics grades, for students in their final year of high school, were related to students' gender, socio-economic status (SES), and scores on a national exam and to school type and location. He documented some findings already reported in the literature, such as a gender gap and an effect of SES, as well as school-type and geographical differences that had not been studied extensively in Italy.

Klapp (2017) took advantage of a natural experiment that was possible because, for a while, Swedish municipalities could decide whether to use grades in sixth grade. Thus, she could compare the 7th-grade GPA of students who had been graded in 6th grade with those who had not (for whom 7th grade was the first year they received grades). She found a negative

effect of grading for low-ability students on 7th-grade GPA, which was fully mediated (that is, the effect disappeared) through academic self-concept in math and Swedish and motivation to improve. She interpreted this to mean that low-ability students with prior experience with grading believed they were poor at math and Swedish and needed to improve, and that they received lower subsequent grades than low-ability students without prior experience with grading.

Studies designed in terms of theory and motivation around learning and engagement, for example, self-regulation and self-confidence, strengthened the theoretical grounding of research on grading, moving from the simple atheoretical studies of yesteryear. One example is the Klapp (2017) study just reported. This trend toward firmer theoretical grounding is also evident in interview and survey studies.

An interview study was conducted in Norway and Sweden with upper-secondary students; despite some differences in grading policies and procedures, grades are high-stakes for students in both countries (Liljeröd et al., 2025) and are an important determinant of admission to higher education. All participants thought grades were important, and most said they studied more because they knew they were being graded. Students saw grades as a public indicator of their competence and, sometimes, a distraction from learning. Good grades provoked positive emotions that enabled students to study harder, and grades also prompted strategic decisions about studying, including sometimes deciding they didn't need to study more. Most students who perceived their grades as low reported negative emotions, reduced willingness to study, and a negative effect on their academic self-concept. A few students reported feeling activated by low grades, taking them as a challenge to show they could do better.

Chen and Bonner (2023) surveyed pre-service and in-service teachers about using student classroom conduct in assigning grades, referencing the learning theories teachers drew from. A number of respondents did support conduct-based grading, despite its being contrary to current recommended practices. Secondary content-area teachers were more likely than elementary teachers to endorse this view. Different teachers held different views, however, and many endorsed achievement-only grading. Teachers who endorsed mixing achievement and conduct in grades pointed out that student conduct affects the learning environment for everyone in the classroom, and they were more likely to endorse traditional or didactic teaching approaches and behaviorist approaches to classroom management.

In my view, studies that add to the theoretical foundation for understanding grading practices are expanding the field in an important way. Of course, there are still some studies of teachers' grading beliefs and practices that situate their investigations in the context of prior research instead of learning theory. Thus, they replicate or extend findings from previous research with other and more contemporary populations. These studies continue to find similar results as in the past (Brookhart et al., 2016). Guskey and Link (2019) found differences in grading practices by grade level; for example, secondary teachers put more weight on formal exams and assignments than do elementary teachers. Kunnath (2017) found that secondary teachers identified concerns about external perceptions, desire to facilitate student success, and administrator pressure for assigning low grades as drivers of their grading practices and mixed achievement and non-achievement factors, especially student ability and effort, in assigning grades. Some survey studies of grading practices replicate or extend previous findings about grading, updated with more recent samples or samples from different countries. For example, Arsyad Arrafii (2020) found that Indonesian secondary school teachers consider student achievement as well as effort and external factors in grading. Moreover, they reported feeling some pressure to inflate grades.

The Pandemic and Research on Grading

Because of the COVID-19 pandemic, most K-12 education was online, at least for a while, in 2020 and 2021. Some students, especially high school students, began to worry about their grades and their futures (Ashta et al., 2023). State departments of education published guidance for grading during emergency remote learning (Townsley & Kunnath, 2022). Troubles in traditional grading practices, many of which SBG had been designed to address, became more obvious in the online environment, like over-reliance on tests and worksheets and inflexible policies that measured work habits as much as learning. For example, some teachers experienced problems grading participation when students had trouble logging in or turned off their video cameras. Tests and quizzes were no longer guaranteed to be done by students individually, without help. Worksheets and other assignments lost context, as teachers could no longer be sure what instructional experiences had actually occurred for each student. Districts experimented with various stopgap grading policies, most of which were not much help.

The pandemic affected grading systems in countries other than the US, as well. For example, in Norway, secondary school students' grades are typically composed of 80% teacher assessments and 20% central exams, and grades determine admission to higher education. When the exams were canceled due to the pandemic, students' final grades rose, which prompted questions about fairness and justice in college admission and both positive and negative consequences for students

(Gamlem et al., 2023). I have no doubt that a search for research about grading during the pandemic in other countries would illustrate, as this does, that the effects of the pandemic interacted with the grading system in place.

At the time of this writing, the COVID-19 pandemic is over. Based on the US experience, Townsley (2020) recommended three principles for pandemic-era learning that are also core SBG principles: base academic grades on prioritized course objectives or standards, report non-cognitive and behavioral conduct separately, and design grading systems that prioritize what students have learned over when they have learned it. Research on SBG principles continues to proliferate. In the next two sections, I provide an overview of some of the advancements in research on SBG implementation and practice and on school leadership for SBG, respectively.

Implementation in SBG

What stands out as new in the K-12 grading research is the proliferation of empirical studies of SBG since 2016. In 2016, we claimed (p. 828), "Literature on SBG implementation recommendations is extensive, but empirical studies are few," and we discussed only seven empirical studies dated from 2003 to 2014. That is certainly not the case today.

When research begins in new topic areas in education, it often begins with descriptive studies investigating practices—what is done—and perceptions—how people think and feel about what is done. In the time since the 2016 grading review article, researchers have begun to publish descriptions of SBG practice and teacher or student beliefs about SBG. Many of these studies used small samples in one or two contexts. Others used larger sample sizes but, because of that, lost the ability to describe the context. The results are mixed, and this could simply be the result of the context-bound nature of this research, with its attendant variation in school and community readiness for SBG, teacher knowledge and skill with SBG practices, and cultural or micro-cultural differences between contexts.

Descriptive Studies of SBG Implementation and Beliefs

The following studies paint a picture of SBG as it emerges into the mainstream. Unlike traditional grading practices where everyone "just knows" what an A means, teachers and students alike are learning about SBG, and halting progress—slow and sometimes hesitant—is being documented. The following five studies provide evidence of this halting progress. I include sample sizes to show that most of them are small-sample, context-bound investigations. While I think this is a good thing, because grading practices like other educational practices are situated in school and community cultures and micro-cultures, it does mean that no sweeping generalizations about SBG reform can yet be made.

Knight and Cooper (2019) studied seven teachers from five high schools with varying levels of SBG in place. They described positive effects for instruction, assessment, the classroom environment, and students' approaches to learning. However, all the teachers needed to convert their standards-based grades to percentages and letter grades for reporting, which essentially meant that what was studied was a standards-based approach to traditional grading. Even so, Knight and Cooper found evidence for an implementation dip, where newly adopted practices get more confusing before they become clear and routine.

Olsen and Buchanan (2019) studied 15 teachers in two high schools who were undergoing a year-long professional development on grading reform. The goal of the professional development was to reform grading practices that included behavior, effort, and motivation in the same academic grade by using separate academic grades. One of the two high schools had already moved to a 1-4 scale, away from percentage grading. Teachers had little prior professional learning on grading and expressed varied and unclear purposes for grading. They also expressed differences in philosophies about the purpose of education in general and, therefore, what elements of student performance they believed were gradable. In these respects, teachers were influenced by their own backgrounds and experiences and by school characteristics. Teachers did make some changes, but they were only partial and sometimes reverted to the grading approach the program was trying to reform.

Ketsman and Reeves (2023) surveyed 136 K-12 teachers in one district participating in SBG reform. They asked questions about SBG practices, teacher perceptions of those practices, perceptions of barriers or challenges to implementation, and teachers' prior opportunities to learn about SBG during pre- and in-service education. Overall, teachers reported mixed levels of implementation of various SBG practices and neutral perceptions of SBG. They tended to agree that SBG provided better information about students' strengths and weaknesses. Barriers identified included time, incompatibility with current grading systems, and parent/guardian perceptions. Teacher confidence partially mediated the relationship between pre-service opportunity to learn and frequency of use of SBG practices, and teacher confidence fully mediated the relationship between in-service opportunity to learn and frequency of use of SBG practices.

Also regarding pre-service preparation, Battistone et al. (2019) investigated whether 11 new teachers were prepared to work in a school that uses SBG. They found that pre-service training programs and field experience were inconsistent in their assessment training.

In one of the larger-scale studies, Wisch et al. (2018) surveyed 429 secondary teachers from 17 schools on their beliefs about SBG practices, especially redo/retake policies. They found that whether teachers allow redos and retakes depended on school policies, content area, and teachers' personal beliefs about the importance of deadlines and student accountability. Many teachers used grading practices that fell somewhere between SBG and traditional grading, although only one of the 17 schools had an explicit SBG policy, and even that school needed to transform its grades to the 100-point scale for reporting. Forty-one percent of the teachers indicated their school had an SBG policy, and 98% said their school allowed redos and retakes. Only 8% of teachers did not allow redos or retakes of any sort; the rest had varying policies but allowed at least some revisions.

The following studies, in music education and physical education, respectively, demonstrate that research on SBG implementation is also arising within content areas. "Halting progress" also seems to describe these efforts.

St. Pierre and Wuttke (2017) conducted a survey of SBG use among music educators in one state. Among the 96 volunteer respondents, just over half reported they were not familiar with SBG. Almost half (47%) reported that they were using SBG at least partially, and they defined what they meant by SBG. After removing some participants whose definitions did not reflect understanding of SBG, the researchers determined that 40% were actually using the practice. The most common reason teachers gave for using SBG was that they were required to, and the most common reason teachers gave for not using SBG was lack of knowledge about it.

Townsley and McNamara (2021) studied implementation of SBG by four physical education teachers and two administrators at one school. They found that while teachers were somewhat confused about what exactly they were supposed to do to implement SBG, the work did result in teachers' desire to more fully understand their state's physical education standards (Townsley & McNamara, 2021). Townsley and McNamara (2024) found that preservice teachers participating in a one-hour SBG workshop helped students appreciate the importance of physical education and grading on defined standards but needed more support before they began implementation in their own classrooms.

The following two studies, one small and one larger but still within one school, show how high school students who have become used to traditional grading are concerned—perhaps with a little of the flavor of "fear of the unknown"—about how SBG will affect their immediate needs, both social and educational.

Guskey et al. (2020) studied students' beliefs about the impact of SBG in high school on their transition to college or university. They used surveys to ascertain the degree of SBG implementation in high school for 13 students, followed by interviews with 7 of those students. They found that high schools were not implementing SBG with fidelity, but to the degree they did implement SBG, students did not perceive these practices had caused problems for their college performance. Concerns about SBG and the transition to college, which some parents voice, may be misplaced; for the students in this study, the social and time management aspects of their transition to college were their biggest concerns.

Peters et al. (2017) studied students' perceptions of SBG at one high school. The school had implemented SBG policies and had worked to help students and parents understand them. However, there was still some "disconnect" (Peters et al., p. 14) for parents and students. They used fall survey data from 376 students and spring surveys from 230 students, out of 478 enrolled in the school. Students had implementation concerns, as not all teachers were consistent in their implementation of the policy. Some students were concerned about the 1-4 scale, worried their grades would be lower, and had fears about their preparation for higher education. Less common were concerns about student complaints, stress, the appearance of being "stupid" if one needed to reassess, and motivation. The authors point out that resistance to change is unsurprising and that specific student concerns can be a source of feedback for the school as it moves forward with SBG implementation.

Finally, the following study gives evidence that communication about SBG, important for any school reform, is important for grading reform, as well.

Larsen (2023) asked 10 teachers, 4 parents, and 2 high school principals to discuss their beliefs about SBG. Ten of the 12 educators believed that SBG improved student learning, even though all educators agreed that letter grading was easier to do. Some parents expressed dismay that homework and other practice work was not recorded, a finding that can be addressed with education and communication.

What are we to conclude from these studies of SBG implementation? The evidence offered in these studies suggests that learning and implementing new grading practices is gradual and influenced by the attitudes and beliefs of the educators involved. As we look at these studies of "halting progress," it is possible to be a pessimist, emphasizing the "haltingly" slow and hesitant nature of changes, mixed results, and the fact that SBG reform is not yet implemented fully and with fidelity to basic SBG principles. It is also possible to be an optimist, emphasizing the "progress" and the fact that over the last

decade these reforms have been pursued and are beginning to be studied, with the intention that future implementation can benefit from understanding the barriers and uncertain beginnings this research has brought to light. A question one might ask is whether SBG reforms are worth it: Is there evidence that SBG is effective? The next section looks at that question, using research that has been published since the 2016 review.

Investigations of the Effectiveness of SBG

For new topics in educational research, efficacy studies often follow descriptive studies when the field has developed to a point where an intervention can be defined clearly enough to evaluate. A few such studies have begun to appear for SBG, investigating the effectiveness of SBG practices on clarity of communication, math anxiety, achievement and homework completion, and achievement and motivation, respectively. Unlike the mixed results of the descriptive studies, these studies all reported positive effects of some kind.

Link and Guskey (2022) reviewed literature to answer the question "Is SBG effective?" They defined SBG as adherence to three principles: (1) report student performance on standards rather than as a single content-area grade; (2) use a limited number of reporting categories; and (3) report academic grades separately from non-cognitive and behavioral evaluations. Interestingly, they did not make multiple opportunities to show mastery, or redos and retakes, one of the defining categories of SBG. Their review supported the conclusion that SBG was an effective communication tool because applying these three principles clarifies the meaning of grades. They did not address effects on students' learning and performance or behavior.

Fergus and Smith (2022) investigated how SBG practices affected five students in two middle-level mathematics classes. These students had self-identified as having math anxiety. Using a case study method, the researchers found that assessment was the characteristic of proficiency-based learning or SBG most closely associated with reducing students' math anxiety. Specifically, establishing clear learning objectives and providing effective feedback were two SBG practices students reported as being helpful.

Huey et al. (2022) studied the two-year effects of removing practice work from grade computations in a geometry course on student performance and behaviors. The students all had the same teacher. The students were primarily high-achieving, and changes in performance due to reassessment were negligible. Two cohorts were compared: for Cohort 1, homework completion counted for 15% of the final grade, and the rest was based on exam scores; for Cohort 2, the final grade was based solely on exam scores. Cohort 1 had higher exam scores and greater homework completion rates. Some students reported seeing homework as optional if it didn't count toward their grade.

Kramer et al. (2024) studied the impact of an SBG program called PARLO (Proficiency-based Assessment and Reassessment of Learning Outcomes) in 9th grade mathematics. This study was a large-scale (n = 2,736) randomized controlled trial, the only such design I have come across so far. Interestingly (to me, anyway), PARLO principles included basing student final grades on the number of learning outcomes for which students achieved proficiency or better, providing opportunities for reassessment for full credit after further learning, using clearly stated learning goals, and providing students with formative feedback. Because the latter two (learning goals and feedback, which are the formative aspects of SBG) were beginning to be widespread and therefore present in both control and treatment (PARLO) schools, the study really evaluated the effects of basing grades on proficiency and using reassessment. The PARLO program (that is, proficiency grades with reassessment) improved student achievement on end-of-course tests. Students with higher expectancies for and value of their mathematics learning benefited more.

One might wonder about these largely positive studies of effects of SBG reform on communication, achievement, motivation, and behavior, given the mixed results or "halting progress" reported in the descriptive studies in the previous section. Is it publication bias? Or is it, perhaps, that the clarity of learning goals and focus on proficiency—by both teachers and students—is so powerful that it's actually helpful even when diluted by less-than-full implementation? I don't think we have enough evidence in the studies reported here to draw a conclusion, but it's worth thinking about.

Two additional studies are reported below, one investigating the effects of SBG on predicting ACT scores and the other on the long-term economic effects of the alternative grading practice of minimum grading. Both are distal, rather than proximal, effects of SBG grading reform, but interesting questions to ponder. The results were not quite as positive as the studies of proximal effects reported just above.

Townsley and Varga (2018) compared students' GPAs from two demographically matched high schools that practiced traditional grading and SBG, respectively. They also compared the ACT scores of these two groups and investigated whether GPA and grading method predicted ACT scores. GPAs did not differ between schools, but students from the traditional-grading school had higher ACT scores (English, Math, and composite). Regarding prediction, English or Math GPA was

the best predictor of ACT English or Math score, respectively; males scored higher than females in each, and students from the traditional-grading school scored higher than those from the SBG school.

Morris and Johnson (2024) simulated the economic impact of implementing a statewide minimum grading policy in ninth grade in Arkansas. Minimum grading is an alternative method used with percentage-based grading that assigns a lower limit to low grades, typically 50 instead of 0, to mitigate the effects of failing grades on grades assigned on the 0-100 scale. The simulation investigated how minimum grading, a no-cost intervention, might increase school effectiveness (by reducing expenses required to teach students who don't graduate), increase the cost-effectiveness of public-school spending, and raise the return on investment in terms of students' lifetime earnings divided by the cost of educating them. They estimated that a 50% minimum grading policy would affect 1 to 3 percent of students (pass at least one additional course), increase educational cost-effectiveness by 6 to 23%, and have no impact on return on investment in terms of students' future earnings.

Leadership in SBG

There is a small group of studies of principals and other educational leaders' perceptions of SBG. These studies are descriptive, similar in tone to the descriptive studies reported above but describing the thoughts and actions of educational leaders instead of teachers. They were all conducted in the same state and by the same team of colleagues, with one exception: a study done with school principals in China (Cheng et al., 2023). These studies in one state do not generalize to the rest of the US; however, I report them here because of this journal's focus on administration and because leadership is one of the classic, well-established, and well-researched foundations that can make or break educational reform. If SBG practices are to reform grading and have the desired effects on students' beliefs, achievement, and motivation, educational leaders will need to know how to proceed.

Buckmiller et al. (2020) surveyed principals of rural high schools in one state. They overwhelmingly reported envisioning implementing SBG practices in the next 5 years, indicated a good understanding of the major ideas behind SBG grading reform, and believed they had the resources to implement. Slightly less, but still a majority, agreed that their teachers would have the disposition to engage in grading reform and that their upper-level administrators and school boards would support SBG reform. They also expected some resistance from parents and community members. Using data from the same survey, Townsley et al. (2019) predicted a second wave of SBG implementation would follow the early adopters or first wave of SBG (Peters & Buckmiller, 2014) in the state, and that principals felt teacher resistance to be the most common barrier. In addition, they believed that SBG grading reform would require time and planning, teacher professional development, and community education.

Townsley and Knight (2020) studied SBG reform in one high school. Their case study described implementation, barriers—including mixed levels of buy-in and depth of understanding of SBG—and the amount of effort required to produce proficiency scales that are well calibrated for cross-classroom use but meaningful enough to use within classrooms.

Research on Rubrics

Another area of study in which empirical literature has expanded greatly in the last decade is rubrics. Research in this topic area is past the stage of purely descriptive studies and has developed to the point that some meta-analyses are beginning to be published. These studies show that, overall, the use of rubrics has positive effects on student achievement and self-regulation of learning (Camargo Salamanca et al., 2024; Panadero et al., 2023). Research on rubrics is not usually linked with research on SBG, but it seems to me that a case can be made.

One important change in contemporary grading practices involves the use of proficiency or mastery scales instead of points and percentages. These scales typically have labels like "Advanced, Proficient, Approaching, and Beginning," or any one of several similar level names. These labels support performance level descriptions used for both formative assessment and grading; in other words, many SBG initiatives rely on rubrics.

To be clear, rubrics for assignments or sets of assignments (e.g., narrative writing, math problem solving) are the kind of rubrics that have been studied in the rubric research literature. Ketsman and Reeves (2023) point out the difference between scoring rubrics for assignments and proficiency scales for evaluating a body of evidence. While SBG does use proficiency scales designed for evaluating a body of evidence, it also quite commonly uses rubrics for assignments coordinated with those proficiency scales.

So far, I have not seen anyone try to combine the SBG and rubric literatures, but it seems obvious to me that the descriptions of proficiency on standards matter, and the research on rubrics is relevant to grading and especially to SBG, which rests on identifying and communicating clear standards. I suggest that readers interested in contemporary grading research should also be encouraged to read contemporary rubrics research.

As an example of a study that is relevant to both topic areas, rubrics and grading, I'll describe a study by Quinn (2020), who, through the use of the internet, was able to design a randomized experiment involving over 1,500 teachers. He used this study to investigate whether well-designed grading practices might help with bias in grading. Student writing samples were created that differed only by cues about the author, which were evident in the names used (distinctively Black or distinctively White). He found that when teachers evaluated a writing sample with a vague scale (e.g., "far below grade level" and so on), teachers scored the same writing sample lower for the Black author than the White author. However, when they used a descriptive rubric for personal narratives, there was no evidence of racial bias in grading.

FUTURE DIRECTIONS FOR GRADING RESEARCH

In addition to continued research about SBG, I see at least three additional areas where grading research may grow, advance, and extend in the future. These are, in no particular order of likelihood or prediction, AI-assisted grading, competency-based grading (and perhaps some other alternative grading practices that haven't been invented yet), and equitable and culturally relevant grading.

Already existing, but mostly so far studied in the context of higher education, is a literature about the possibilities and benefits of AI-assisted grading (e.g., Flodén, 2025). I expect that to continue, maybe even explode, and to extend further into the K-12 context. The interest is in AI grading as a labor-saving device for teachers, which is important enough in its own right. However, my own view is that if AI can be harnessed to assist with formative feedback, as a labor-saving device for teachers, that would be even more important. Presumably, regarding assessment for summative purposes, teachers will grade what they must anyway, and as the SBG literature shows us, more grades don't mean better grading. However, regarding assessment for formative purposes and most notably feedback, teachers might give more feedback if it were less burdensome to do so, to the benefit of students and their learning.

Competency-based instruction and assessment have developed greatly since 2016, and in my view, it finds itself where SBG was in 2016: there is a literature of recommendations for implementation, but not much in the way of empirical studies. At present, every state in the U.S. allows for competency-based learning, but it is not yet a reality in schools (Stanford, 2024). I predict that, as for the wave of SBG research before it, there will be a wave of research into competency-based grading—at least, if it is to survive as a practice. Typically, as with SBG research, that would mean a wave of descriptive studies first (to identify what practices are being used in what ways and the perceptions of educators and students about these practices) followed by efficacy studies (to investigate effects on learning and motivation) and more theoretically grounded studies (to investigate the reasons for the effects).

Finally, and importantly, in the decade since 2016, there has been a growing interest in grading practices that are equitable and culturally relevant (Brookhart & Chism, 2025). As for competency-based instruction now (and for SBG ten years ago), there is a large and growing literature on recommendations for implementation of equitable and culturally relevant (e.g., Feldman, 2024) grading practices, but not much yet in the way of empirical studies. I would expect research in this area to proceed in the same manner: descriptive studies, followed by efficacy studies and more theoretically grounded studies.

In summary, research in grading is moving along nicely. I am particularly happy to see development in several topic areas at once that are somewhat different from the traditional grading research we summarized ten years ago (Brookhart et al., 2016). The next 100 years of grading research should be very exciting to researchers and, I hope, consequential for the improvement of practice.

REFERENCES

Arsyad Arrafii, M. (2020). Grades and grade inflation: Exploring teachers' grading practices in Indonesian EFL secondary school classrooms. *Pedagogy, Culture and Society, 28*(3), 477-499. https://doi.org/10.1080/14681366.2019.1663246

Ashta, J. K., Weingart, R., & Gazmararian, J.A. (2023). The impact of COVID-19 on education experiences of high school students in semi-rural Georgia. *Journal of School Health, 93*(4), 257-265. https://doi.org/10.1111/josh.13269

Battistone, W., Buckmiller, T., & Peters, R. (2019). Assessing assessment literacy: Are new teachers prepared to assume jobs in school districts engaging in grading and assessment reform efforts? *Studies in Educational Evaluation, 62*, 10-17. https://doi.org/10.1016/j.stueduc.2019.04.009

Brookhart, S. M., & Chism, D. (2025). Equitable and culturally relevant grading practices. In C. Evans & C. Taylor (Eds.), *Culturally responsive assessment in classroom and large-scale contexts: Theory, research, and practice* (pp. 306-323). Routledge.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4)*,* 803-848. https://doi.org/10.3102/0034654316672069

Buckmiller, T., Townsley, M., & Cooper, R. (2020). Rural high school principals and the challenge of standards-based grading. *Theory and Practice in Rural Education*, 10(1), 92-102. https://doi.org/10.3776/tpre.2020.v10n1p92-102

Camargo Salamanca, S., Parra-Martinez, A., Chang, A., Maeda, Y., & Traynor, A. (2024). The effect of scoring rubrics use on self-efficacy and self-regulation. *Educational Psychology Review, 36,* 70. https://doi.org/10.1007/s10648-024-09906-w

Chen, P., & Bonner, S. (2023). Teachers' beliefs about grading, grades, and student classroom conduct. *Educational Practice and Theory, 45*(2), 69-91. https://doi.org/10.7459/ept/45.2.06

Cheng, L., Baidoo-Anu, D., & DeLuca, C. (2023). Examining school principals' conceptions of assessment and grading practices. *Alberta Journal of Educational Research, 69*(1), 41-65. https://journalhosting.ucalgary.ca/index.php/ajer/

Doz, D. (2023). Factors influencing teachers' grading standards in mathematics. *Oxford Review of Education, 49*(6), 819–837. https://doi.org/10.1080/03054985.2023.2185217

Feldman, J. (2024). *Grading for equity* (2nd ed.). Corwin.

Fergus, S., & Smith, C. P. (2022). Characteristics of proficiency-based learning and their impacts on math anxiety in the middle grades. *Research in Middle Level Education Online,* 45(4), 1-19. https://doi.org/10.1080/19404476.2022.2045810

Flodén, J. (2025). Grading exams using large language models: A comparison between human and AI grading of exams in higher education using ChatGPT. *British Educational Research Journal, 51*(1), 201-224. https://doi.org/10.1002/berj.4069

Gamlem, S., Smith, K., & Sandvik, L. V. (2023). Stakeholder opinions about canceling exams in Norwegian upper secondary school during the pandemic, and its consequences—An illuminative study. *Assessment Matters, 17*. https://doi.org/10.18296/am.0063

Guskey, T. R., & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades, *Assessment in Education: Principles, Policy & Practice, 26*(3), 303-320. https://doi.org/10.1080/0969594X.2018.1555515

Guskey, T. R., Townsley, M., & Buckmiller, T. M. (2020). The impact of standards-based learning: Tracking high school students' transition to the university. *NASSP Bulletin, 104*(4), 257-269.

Huey, M. E., Silvey, P. R., Vaughan, A. G., & Fisher, A. L. (2022). Assessing the impact of standards-based grading policy changes on student performance and practice work completion in secondary mathematics. *Studies in Educational Evaluation, 75,* 101211. https://doi.org/10.1016/j.stueduc.2022.101211

Ketsman, O., & Reeves, T. (2023). Standards-based grading in a small, suburban district: Teacher education, confidence, and implementation. *Journal of School Administration Research & Development, 8*(2), 65-75. https://doi.org/10.32674/jsard.v8i2.5565

Klapp, A. (2017). Does academic and social self-concept and motivation explain the effect of grading on students' achievement? *European Journal of the Psychology of Education.* https://doi.org/10.1007/s10212-017-0331-3

Knight, M., & Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin, 103*(1), 65–92. https://doi.org/10.1177/0192636519826709

Kramer, S. L., Posner, M. A., Browman, A. S., Lawrence, N. R., Roem, J., & Krier, K. (2024). The impacts of a standards-based grading system emphasizing formative assessment, feedback, and re-assessment: A mixed methods, cluster randomized control trial in ninth grade mathematics classrooms. *Journal of Research on Educational Effectiveness, 18*(1). https://doi.org/10.1080/19345747.2023.2287594

Kunnath, J. P. (2017). Teacher grading decisions: Influences, rationale, and practices. *American Secondary Education, 45*(3), 68–88.

Larsen, A. (2023). Investigating the potential benefits of standards-based grading practices at urban secondary schools in Southern Utah: A qualitative study. *Journal of Nonprofit Innovation,3*. https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1076&context=joni

Liljeröd, H., Jönsson, A., Klapp, A., & Jonsson, A.-C. (2025). Students' perceptions of how grades influence their motivation: Voices of upper secondary school students in Norway and Sweden. *Educational Assessment, Evaluation and Accountability,* https://doi.org/10.1007/s11092-025-09454-z

Link, L. J., & Guskey, T. R. (2022). Is standards-based grading effective? *Theory into Practice, 61*(4), 406-417. https://doi.org/10.1080/00405841.2022.2107338

Morris, S. R., & Johnson, A. H. (2024). Could minimum grading enhance high school graduation rates and cost-effectiveness across Arkansas? *Education Policy Analysis Archives, 32*(40). https://doi.org/10.14507/epaa.32.8496

Olsen, B., & Buchanan, R. (2019). An investigation of teachers encouraged to reform grading practices in secondary schools. *American Educational Research Journal, 56*(5), 2004-2039. https://doi.org/10.3102/0002831219841349

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: A meta-analytic review. *Educational Psychology Review, 35*, 113. https://doi.org/10.1007/s10648-023-09823-4

Peters, R., & Buckmiller, T. (2014). Our grades were broken: Overcoming barriers and challenges to implementing standards-based grading. *Journal of Educational Leadership in Action, 2*(2), Article 2. https://digitalcommons.lindenwood.edu/ela/vol2/iss2/2

Peters, R., Kruse, J., Buckmiller, T., & Townsley, M. (2017). "It's just not fair!" Making sense of secondary students' resistance to a standards-based grading initiative in the Midwestern United States. *American Secondary Education, 45(*3), 9-28.

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis, 42*(3), 375-392. https://doi.org/10.3102/0162373720932188

Stanford, L. (2024, September 17). All states allow competency-based learning. Will it become a reality in schools? *Education Week.* All States Allow Competency-Based Learning. Will It Become a Reality in Schools? (edweek.org)

St. Pierre, N. A., & Wuttke, B. C. (2017). Standards-based grading practices among practicing music educators: Prevalence and rationale. *Update: Applications of Research in Music Education, 35*(2), 30-37. https://doi.org/10.1177/8755123315604468

Townsley, M. (2020). Grading principles in pandemic-era learning: Recommendations and implications for secondary school leaders. *Journal of School Administration Research and Development, 5*(S1), 8-14. https://doi.org/10.32674/jsard.v5iS1.2760

Townsley, M., Buckmiller, T., & Cooper, R. (2019). Anticipating a second wave of standards-based grading implementation and understanding the potential barriers: Perceptions of high school principals. *NASSP Bulletin, 103*(4), 281-299. https://doi.org/10.1177/0192636519882084

Townsley, M., & Knight, M. (2020). Making change stick: A case study of one high school's journey towards standards-based grading. *Journal of Educational Leadership in Action, 6*(3), Article 3. https://doi.org/10.62608/2164-1102.1014

Townsley, M., & Kunnath, J. (2022). Exploring state department of education grading guidance during COVID-19: A model for future emergency remote learning. *Education Policy Analysis Archives, 30*(163). https://doi.org/10.14507/epaa.30.7448

Townsley, M., & McNamara, S. (2021). "I thought I was supposed to get an A in PE!" Successes and challenges of teachers and administrators implementing standards-based grading in physical education. *Studies in Educational Evaluation, 70.* https://doi.org/10.1016/j.stueduc.2021.101012

Townsley, M., & McNamara, S. (2024). "PE is more than playing kick ball the entire hour": Using standards-based grading to elevate physical education's status among preservice physical educators. *Journal of Teaching in Physical Education*, *43*(4), 733-739. https://doi.org/10.1123/jtpe.2023-0064

Townsley, M., & Varga, M. (2018). Getting high school students ready for college: A quantitative study of standards-based grading practices. *Journal of Research in Education, 28*(1), 92-112. https://eric.ed.gov/?id=EJ1168171

Wisch, J. K., Ousterhout, B. H., Carter, V., & Orr, B. (2018). The grading gradient: Teacher motivations for varied redo and retake policies. *Studies in Educational Evaluation, 58*, 145-155. https://doi.org/10.1016/j.stueduc.2018.06.005

SUSAN BROOKHART, PhD, is a Professor Emerita in the Department of Educational Foundations and Leadership at Duquesne University. Her major research interests are the role of both formative and summative classroom assessment in student motivation and achievement, the connection between classroom assessment and large-scale assessment, and grading. Email: suebrookhart@gmail.com

Manuscript submitted: ***June 26, 2025***
Accepted for publication: ***October 30, 2025***

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 63-73
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

The 5 Ps of Starting and Sustaining Grading Reform

Matt Townsley
University of Northern Iowa, USA
Joshua Kunnath
Highland High School, Bakersfield, CA, USA

ABSTRACT

In the world of grading reform, questions such as "How do we get started with grading reform?" and "Our teachers completed a book study—what's the next step?" are at the forefront of school leaders' minds. Engaging conference keynote addresses and multi-day seminars can help K-12 educators understand the need to improve grading practices; however, school leaders may need practical steps to start or sustain grading reform. These practical steps should be based upon current research in the field and supported through a cohesive set of actions. This article presents a 5-P Framework for starting and sustaining grading reform.

***Keywords*:** grading reform, standards-based grading, educational leadership, implementation framework, change management

A strong body of research suggests that schools should consider grading practices that are more consistent (Brookhart & Guskey, 2019), more equitable (Brookhart & Chism, 2025; Griffin & Townsley, 2022), and that better communicate learning (Brookhart et al., 2016). Grading experts recommend that schools transition away from points, percentages, and the mixing of achievement and non-achievement factors (Guskey, 2020a; Link & Guskey, 2022; Townsley, 2022). However, school leaders' existing knowledge, mindsets, and experiences may constrain their readiness to initiate or sustain meaningful grading reform (Buckmiller et al., 2020; Kunnath et al., 2026; Morris et al., 2025). As a result, a "knowing–doing gap" (Pfeffer & Sutton, 2000) can emerge, in which leaders understand the need for change but struggle to translate that knowledge into action. Furthermore, research in the field suggests that initiating grading reform or expanding effective grading practices presents challenges that school leaders may not be adequately prepared to address (Peters & Buckmiller, 2014; Townsley et al., 2019).

For example, one of the authors of this article received a phone call from a Kansas high school administrator whose building leadership team was energized to begin implementing more effective grading practices. Classroom teachers on the building leadership team had attended multiple conferences focused on grading and had recently participated in a grading-focused book study; however, the administrator was unsure of the steps to initiate effective grading practices at their school. During the next 18 months, this author, the high school administrator, and the building leadership team enacted a 5-P framework to not only initiate grading reform but also build capacity within the system to sustain these important changes beyond initial implementation. Utilizing this step-by-step framework, classroom teachers who were members of the building leadership team served as a guiding coalition to unify their understanding of effective grading practices, establish agreed-upon grading policies, and strategically implement these policies in their classrooms. Through their enhanced learning, feedback loops, and firsthand experiences, the building leadership team later felt confident enough to support their colleagues in implementing these grading policies. Today, this high school in Kansas continues to implement standards-based grading practices despite administrative turnover and serves as an example of the importance of what Townsley and colleagues (2019) recommend: "go slow to go fast" (p. 291).

In this article, we present a pragmatic approach to starting and sustaining grading reform with a 5-P Framework consisting of people, purpose, policies, pace, and persistence. Our framework begins with the right people, and then these people develop a deep understanding of *why* change is needed, followed by a multi-step process for getting started and sustaining grading reform. The 5 Ps are previewed in Figure 1 below.

Figure 1

A Five P Framework for Starting and Sustaining Grading Reform

People	Assembling the right people to lead and facilitate grading reform.
Purpose	Establishing the *why* and *what* of grading reform by co-creating a grading purpose statement.
Policies	Communicating the *how* of grading reform through identifying a list of 3-5 grading policies in support of the grading purpose statement.
Pace	Intentionally honoring the complexity of grading reform by implementing the grading policies over time.
Persistence	Providing sustained support to implementation by utilizing feedback loops aligned with the established grading policies.

PEOPLE

The first P of the framework is getting the *right people* in place to spearhead the work. Instead of starting with grading policies or even a vision, we start here because it's the people who lead and perform the work, not the policies or purpose statement. Jim Collins (2001), author of the highly influential business management book *Good to Great*, explored what separates consistently great companies from those that are merely good, devoting his third chapter to finding the right people. He found that "executives who ignited the transformations from good to great did not first figure out where to drive the bus and then get people to take it there." Collins clarified that rather than starting with purpose and vision, these leaders "*first* got the right people on the bus (and the wrong people off the bus) and *then* figured out where to drive it" (p. 41). Similarly, we suggest that a school leader's first step in grading reform is to identify the right people to lead the change they hope to enact in their school.

Without a team of the right people, the effectiveness and longevity of grading reform are limited to the time, effort, and tenure of the person pushing the initiative, typically a site or district-level administrator. Furthermore, effective leadership in grading reform requires a complex combination of vision, passion, experience, expertise, credibility, and trust. And while a single person is highly unlikely to possess these qualities, a thoughtfully selected team can do so collectively. Both John Kotter, Harvard Business School professor emeritus and author of *Leading Change* (2012), and James Spillane, Northwestern professor and author of *Distributed Leadership* (2006), dispel the myth of the "heroic leader" as the primary driver of organizational change in their influential leadership texts. Spillane (2006) replaces it with the concept of distributed leadership from a "leader-plus" perspective, which encompasses multiple individuals across a school site, including administrators, department chairs, instructional coaches, teacher-leaders, counselors, and other respected faculty members. Collectively, these people become the faces and voices of the reform movement.

But assembling such a group requires intentionality, coordination, and structure. Kotter (2012) first coined the term "guiding coalition," calling on a group of individuals within organizations with credibility, expertise, leadership skills, and positional power to drive reform efforts. Such a group, according to Kotter, shares a common commitment, builds trust, communicates effectively, and models the change it desires. These ideas have been championed and applied to schools and, in particular, to professional learning communities by Dufour et al. (2010) and, more recently, by Hall (2021). And although the group has no "heroic leader," it does require an "initiator," usually an administrator with positional power and the urgency to act, to select and convene the group in the first place (Guskey, 2020b).

For this reason, we suggest forming a leadership group directly responsible for grading reform. Thus, a *grading guiding coalition* is needed to move grading reform forward when the scope of the work is large—whether it's a cohort of early adopters, a school site, or an entire district. The coalition ideally includes a strategic mix of stakeholders directly involved

in the grading process, including teachers, instructional coaches, support staff, counselors, and administrators. Such a composition ensures a diversity of credibility and perspective, and with at least one administrator on board, the group has the authority to take real action. Although some members have no positional authority, they lead by example and, in doing so, show other staff members that grading reform is possible at their school site and with their students. Guskey (2020b) also recommends including parents and students—either as direct members or by capturing their voices and perspectives through surveys or focus groups. While the group likely has many differences, they should share at least one thing in common: they find the topic of grading reform to be compelling.

The total size of the coalition can vary greatly, but Kotter (2012) provides helpful guidance, noting that "change often starts with just two or three people," but grows from "half a dozen" to "twenty to fifty," depending on the size of the organization (p. 61). Once convened, the group should establish a common knowledge base on the problems with traditional grading and methods of grading reform. See Appendix A for a list of recommended books and articles to serve this purpose. Next, the work of the grading guiding coalition truly begins with the framework's second P.

PURPOSE

With the right people in place, the leadership team can move to the second P: *purpose*. In this framework, purpose includes two dimensions: *the why* and *the what*, as summarized in Figure 2. Collectively, they constitute a vision for the work—something Kotter (2012) describes as essential for change. *The why* dimension represents the values and beliefs that give the work meaning and urgency. It is the pathos of purpose. Rather than diving in by determining the particulars—the percentage of weights assigned to formative and summative assessment in the grade book or whether the school should permit retakes—a strategic start begins with a broad-based purpose: *the why*. When the guiding coalition begins by establishing their *why* for grading reform, they determine shared understandings and values that provide the determination and drive to take on the inevitable challenges and unknowns that arise when addressing *the how* of grading reform. This collective urgency helps drive future action and is part of the team's essential trust-building process (Kotter, 2012). Simon Sinek (2009), in his book *Start with Why: How Great Leaders Inspire Everyone to Take Action*, points out that it's *the why* that ultimately motivates people to act. For these reasons, we recommend beginning with *the why* of purpose before proceeding to *the what*.

Establishing a common *why* for grading reform requires both reflection and discussion. Here are two guiding questions to facilitate the process:

1. *How are grades important in our students' current and future lives?*
2. *What are the possible implications of ineffective grading practices?*

Once the team has had time to reflect and discuss, they can identify common ideas to drive their work. Below is an example of a *why statement* that incorporates values and beliefs about integrity, decision-making, and equity that may emerge from the questions above:

> *The Central High Grading Guiding Coalition asserts that effective grading practices are essential for creating meaningful report card grades that support valid decision-making regarding student opportunities and areas for growth and development. We further recognize that we must make a deliberate effort to ensure that such practices and grades are applied consistently across all student groups at our school.*

Such a statement may appear at the top of coalition agendas and serves as a continual reminder about the overarching meaning that drives their work.

Furthermore, the "why level" purpose must be established at two levels—the guiding coalition level and, eventually, the broader site level. Because if a school is to successfully implement grading reform at scale, a majority of teachers must buy into *the why*. And involving them in creating a common *why* can be much more effective than simply providing it to them (Deci & Ryan, 2000; Harris et al., 2022; Spillane, 2006). So we suggest repeating the process outlined above, but this time including the entire staff.

Figure 2

Two Dimensions of Grading Purpose

Dimension	Meaning	Function
The *Why*	moral and educational impetus for grading reform	Provide inspiration and drive the work: the pathos of purpose
The *What*	function and meaning of report card grades	Focus grading practices, reporting methods, and decision-making: the logos of purpose

The corollary to *the why* is *the what* dimension. Here, educators establish a clear, common understanding of the purpose that report card grades are intended to serve and the meaning they are intended to convey. By doing so, teachers can ensure they use grading and reporting practices as intentional means that achieve these established ends. Especially important is the *common* element of *the what* dimension, so that all students' grades have the same meaning and function. When this is established and communicated across a school site, it can increase integrity in the grading process, as well as clarity and trust among students and parents (Guskey et al., 2024). It is important to note that, on the whole, this has not been true in American schools for more than 100 years (Brookhart, 2026; Brookhart et al., 2016).

A great deal of work has already been done on this dimension of the grading purpose—particularly by grading experts Susan Brookhart (2009, 2011, 2017) and Thomas Guskey (2015, 2020b, 2024). While their ideas vary in nuance, they are parallel in their emphasis that the main purpose of teacher grading should be to communicate student academic achievement, and the primary audience should be students and parents. We echo these recommendations as both researchers and practitioners. And as a part of the grading guiding coalition's common knowledge-building process, it is important that the group establish an internal *what* of purpose that closely aligns with these expert recommendations. This ensures that the leadership group has a "true north" for the grading reform movement that is supported by leading experts in the field (Schimmer, 2016).

But like the *why dimension*, providing expert recommendations is not a substitute for leading the staff in a collaborative process of discussing and establishing their own *what* for grading purpose. Rather, such recommendations can serve as guidance and even "guardrails," but the teachers themselves should have the final say. When starting the work, the grading guiding coalition may choose to front-load expert recommendations alongside the common *why* the staff previously created. From here, a focused staff discussion can help to organize ideas and identify commonalities. Below are three discussion questions that can help the staff identify *the what* of grading purpose:

1. *What purpose do we intend our report card grades to be used for?*
2. *What common meaning do we assign to our report card grades?*
3. *Who are our report card grades primarily intended for?*

Although answers will usually vary, the coalition members guiding the work are looking for patterns in teacher responses that align with the ideas in both the expert guidance and the staff's *why statement*. Making decisions can be challenging, but the goal is consensus about what Brookhart (2011) calls an "'I can live with that' status" (p. 12). A final step in the creation process is packaging the most important common ideas that emerge from this collaborative process, which can help to guide grading reform throughout the next three Ps of the 5P framework. Figure 3 displays a template that schools can use to establish their common purpose statement.

Figure 3

Grading Purpose Statement Template

The teachers of [team or site] intend for our report card grades to be primarily used to [state or list purpose]. The report card grades we assign are meant to represent [state meaning of grades]. The primary audience for these grades are [state audience].

POLICIES

The third P in the framework is to create a list of three to five grading policies describing the "how" of grading reform. Following *purpose*, Guskey (2020b) suggests educators are now ready to make decisions about the best way to implement grading reform. Depending upon the size of the guiding coalition, the policy work may be accomplished by a smaller group within the guiding coalition or as a whole. Regardless of the size of the team leading this work, the list of three to five policies supports the grading purpose statement and establishes clarity around the "how" of grading reform. Oftentimes, a school or district captures these grading policies within a building handbook, such as a grading practices implementation guide or embedded within other district handbooks or guidelines. Common grading policies focus on five areas of implementation: formative vs summative evidence, number of performance levels, communicating non-achievement separately, reassessment, and extra credit.

Formative vs. Summative Evidence

Experts recommend that grades should be based solely upon summative evidence, rather than a mix of formative and summative evidence (Brookhart, 2017; Heritage, 2022). Historically, grades have been based upon point accumulation regardless of purpose (Townsley, 2022). In contrast, assessment literacy principles indicate that the purpose of formative assessment is to provide students with a clear understanding of where they are at *during* the learning process (Heritage, 2022). Because instruction is ongoing and learning continues after formative assessment activities, it does not make sense to include evidence from formative assessment when determining a student's grade. Similarly, homework and other in-class assignments or activities become feedback and formative in nature, and therefore shouldn't contribute to a student's grade (Townsley, 2022; Vatterott, 2015). Grading policies should emphasize that evidence from summative assessments such as end-of-unit tests, projects, and essays should determine academic grades (Guskey et al., 2024; Townsley, 2022). As a corollary, all other non-summative assignments and assessments will be feedback-only. A sample grading policy is "Summative assessments will be used to determine the current level of learning for each student."

Number of Performance Levels

The results from common standardized assessments, such as state assessments, communicate student learning using a small number of defined performance levels, such as *Standard Not Met*, *Standard Nearly Met, Standard Met,* and *Standard Exceeded*. Although we do not recommend the specific naming conventions for these performance levels (more on this below), classroom teachers should report learning using a finite number of performance levels (Guskey, 2024; Townsley & Wilcox, 2024). Within traditional grading systems, learning is often communicated through points that are converted to percentages. Through what Guskey (2024) refers to as the "illusion of data validity," a percentage-based, 0-100 system is commonly believed to be more precise, given it has 101 performance levels. That is, a student with a 95% on a test may believe they have a higher level understanding than a classmate with a 94% on the same test; however, Guskey (2024) warns that inter-rater reliability and other interpretation errors can make directly measuring student learning down to a percentage point a nearly impossible endeavor. Instead, grading reform authors and educational measurement experts agree that educators report learning using a finite number of performance levels to improve accuracy (Brookhart, 2013; Guskey, 2024; Townsley & Wilcox, 2024). Grading policies should require that learning will be communicated on a report card or within an electronic grade book using a finite number of performance levels—Guskey (2024) recommends four to seven. Sample classroom performance levels may include *Beginning, Developing, Nearly Meets, Meets Standard*[1]. When naming the top level, Schimmer (2016) and Guskey (2014) warn educators that using labels such as *Exceeds Standard* can present implementation difficulties. Their rationale is that when students are required to exceed the standard to reach the top performance level, teacher expectations can be difficult to clarify and defend. A sample grading policy is "The grade book will reflect student learning based upon state standards using four agreed-upon performance levels."

[1] When needed for the purpose of determining course letter grades, each performance level can also have a number connected to it (e.g., 1-*Beginning*, 2-*Developing*, etc.). For a detailed explanation of options for determining letter grades from these levels and corresponding numbers, for example in a high school setting, see Townsley and Wear (2020) or O'Connor (2022).

Communicate Non-achievement Separately

Evidence from over a century of grading research suggests that factors contributing to classroom grades have included a combination of achievement and behavior factors (Brookhart et al., 2016). When behaviors such as participation, responsibility, and attendance are combined with academic achievement, the result is an overall omnibus grade that does not serve a clear purpose. Instead, experts recommend communicating non-achievement separately from achievement (Link & Guskey, 2022; Schimmer, 2023; Wilcox & Townsley, 2022). At the elementary or middle school level, communicating non-achievement separately may look like a "work habits" or "employability skills" section on the report card that includes a small number of habits or skills. At the high school level, communicating non-achievement factors separately may involve ongoing communication using district-provided tools such as Parent Square or an "employability" section within the electronic gradebook. Regardless of the method used to separately communicate these non-achievement behavior factors, educators should seek to ensure these factors are explicitly taught and formatively assessed prior to summative reporting (Schimmer, 2023). A sample grading policy that a guiding coalition may consider related to communicating non-achievement separately is as follows: "Non-achievement factors will be explicitly taught, assessed, and communicated separately from academic achievement."

Reassessment

Educators should provide intentional opportunities for students to reassess their learning (Townsley & Wilcox, 2024). Providing intentional reassessment opportunities was a prominent change during COVID-era emergency remote learning (Townsley & Kunnath, 2022) and appears to be continuing in the years following the pandemic (Kunnath et al., 2026). Although Link and Guskey (2022) suggest that reassessment is not a critical component of effective grading practices, other experts have argued that reassessment is a cornerstone of grading reform (Townsley & Wilcox, 2024). Providing students with opportunities to reassess their learning requires a philosophical and a technical shift in practice. Philosophically, reassessment asks teachers to embrace a learning-centered mindset, recognizing that students learn at different rates and that initial attempts likely involve errors and misconceptions.

However, providing students multiple opportunities over time to demonstrate their learning must be thoughtfully structured so that it does not become a "blind" reassessment, in which policy permits students to complete a second assessment, after performing poorly on the first, without first demonstrating any new learning (Townsley & Wear, 2020). Unlimited, unstructured retakes can unintentionally reduce student accountability and create unsustainable workloads for teachers. Intentional reassessment policies balance opportunity with responsibility by establishing clear expectations and protocols. Time limits or reassessment windows, for example, help ensure that reassessment remains meaningful for students while feasible for teachers. Grading reform authors contend that prerequisite corrective activities, such as error analysis, targeted practice, conferencing, or additional instruction, are equally important (Guskey, 2023; Townsley & Wear, 2020). These steps communicate that improvement comes from relearning rather than blind repetition. A sample grading policy related to reassessment is as follows: "Students will have multiple opportunities to demonstrate learning, with reassessment permitted only after completing an intentional relearning plan."

Extra Credit

Teachers should eliminate the use of extra credit (O'Connor et al., 2018; Reeves, 2016). Too often, extra credit is awarded for "filler" tasks, such as completing holiday crossword puzzles, or for non-academic incentives like bringing supplies such as markers and tissues to school for classroom use. O'Connor and colleagues (2018) do not mince words regarding the use of extra credit when they prescribe, "Leaders must require the elimination of common practices that compromise the accuracy of grades. These include…extra credit that awards points for behaviors that are unrelated to the standards" (p. 69). Given this support for eliminating extra credit, a sample grading policy is "Extra credit will not be given at any time."

All five sample grading policies, which focus on formative vs summative evidence, number of performance levels, non-achievement factors, reassessment, and extra credit, are shown in Figure 4 below. Because context matters and all grading policies should be developed in support of a grading purpose statement, we encourage guiding coalitions to consider Figure 4 as a reference, rather than a model to be utilized verbatim.

Figure 4

Sample Grading Policies

Central Junior High Grading Policies
1. Summative assessments will be used to determine the current level of learning for each student. 2. The grade book will reflect student learning based upon state standards, using four agreed-upon performance levels 3. Non-achievement factors will be explicitly taught, assessed, and communicated separately 4. Students will have multiple opportunities to demonstrate learning, with reassessment permitted only after completing an intentional relearning plan. 5. Extra credit will not be given at any time.

PACE

The fourth P is pace. A common temptation in grading reform is to go "all-in" by setting a firm start date and expecting every teacher to fully implement the grading policies at once. While this approach may appear decisive, change leadership research suggests it is rarely the most effective or sustainable path. For example, Everett Rogers' (2003) diffusion of innovations theory emphasizes the importance of early adopters and gradual spread, while Kotter's (2012) change framework highlights the value of building momentum through short-term wins. Together, these perspectives point to a more intentional and phased approach to implementation. Rather than mandating immediate, comprehensive changes to grading practices, school leaders can begin by identifying small, manageable entry points that allow educators to experiment, reflect, and refine their practice. One example is to incrementally adjust the balance between formative and summative assessment. A school might begin with a 70/30 summative-to-formative weighting and then increase the summative percentage by 10% each year toward an eventual 100% summative weight. This allows teachers, students, and families to adapt to the shift while keeping grading practices stable enough to maintain trust.

Additional strategies include initiating and expanding implementation at the grade level or by content teams. For example, the 6th-grade team may be ready to begin adopting the grading policies in year one, followed by the 7th-grade team in year two. Alternatively, the science team may implement the grading policies in the first semester and then support the math team as they implement the policies in the second semester of the same school year. Regardless of the pacing strategy, these early adopters should receive targeted support (see "Persistence" below), develop shared language and tools, and generate concrete examples of success. Their work creates credibility and practical models for other educators in the school while reducing long-term anxiety across the system. By pacing implementation thoughtfully, leaders honor the complexity of grading reform and increase the likelihood that changes will sustain rather than fade over time. Sustaining grading reform over time also includes supporting teachers throughout various implementation phases, which leads to the next P, persistence.

PERSISTENCE

Persistence is the final P in the framework, which we define as providing sustained support for educators over time. As reformed grading practices are piloted intentionally in classrooms, establishing meaningful feedback loops aligned with established grading policies becomes essential to ensure implementation proceeds intentionally rather than haphazardly. Early adopters within the pilot group often lead this work, but their success depends on sustained support rather than one-time training sessions. Previous research suggests that educators seeking to implement effective grading practices require multiple iterations of support over time (Townsley & Knight, 2020; Townsley & Sult, 2026).

Similarly, research on organizational change within schools consistently shows that meaningful instructional shifts take multiple years to fully take root and require coherent, ongoing support systems (Hall & Hord, 2020; McLure & Aldridge, 2023). This reality underscores the importance of designing implementation plans that include planned opportunities for feedback, reflection, and recalibration. Anecdotal evidence from cohort-based models at one of the author's schools reinforces this conclusion. Within these grading reform cohorts, teachers have engaged in sustained collaborative learning, iterative practice, and action research, rather than rapid, top-down policy adoption. These feedback loops allowed small adjustments to compound into meaningful progress.

Intentional feedback loops should center on specific grading policies rather than the grading reform as a whole. For example, school leaders may choose to establish clear criteria for reassessment (e.g., a written policy that includes corrective activities completed within an agreed-upon timeline) and ask teachers to submit their reassessment policies to school leaders for feedback. Using the results of this feedback loop, teachers may be affirmed in their reassessment details, asked to revise and resubmit for additional feedback, or participate in follow-up reassessment professional learning. Another example of an intentional feedback loop could be multi-faceted. First, early implementing teachers provide later implementing teachers with feedback on their standards-based grade books one month following whole-staff professional learning on setting up their electronic grade books. Using the results of this feedback loop, teachers revise their grade books. Second, to ensure more consistent practices across the building or district, administrators provide feedback to all teachers two months following whole-staff professional learning on their grade books.

Probing implementation through intentional feedback loops signals that progress matters more than perfection and creates a culture of professional growth. When leaders listen closely, respond thoughtfully, and adjust support accordingly, teachers are more likely to persist through the complexities of grading reform, and progress is more likely to be sustained in the long term.

CONCLUSION

Impactful and enduring grading reform requires not only knowledge on the topic, but also deliberate planning and execution. The 5-P Framework, comprising *people*, *purpose*, *policies*, *pace*, and *persistence*, is a strategic reform method designed to help school leaders close the all-too-common knowing-doing gap to put grading theory into practice. Using the framework, the initiating leader begins by getting the right people in place to perform the work by organizing a grading guiding coalition. These people then establish a grading purpose by defining the *why* and *what* of grading, which guides the work of the coalition and individual teachers who implement it in their classrooms. Subsequently, the grading guiding coalition is ready to determine the *how* of reform by deciding on three to five grading policies, considering high-leverage topics such as formative vs summative evidence, number of performance levels, communicating non-achievement factors separately, reassessment, and eliminating extra credit. And while planning for the implementation of these policies, reformers are intentional and deliberate about their pacing, avoiding the common mistake of doing *too* much *too* soon. Finally, the coalition sets the stage for long-lasting reform by establishing feedback loops that will enable teachers and leaders to engage in an iterative process of implementation, feedback, reflection, and refinement. See Appendix A for a list of recommended readings to establish a common knowledge base on grading reform specific to the first three Ps. (Due to the contextual and granular nature of pace and persistence, this appendix does not include resources for these two Ps.)

Ultimately, the goal of grading reform is for such refined grading policies to become the new normal—part of the school's pedagogical culture. This means that reformed grading becomes institutionalized and, regardless of leadership or faculty changes, becomes part of the organizational professional practice. The 5-P framework was a key to implementing successful grading reform for the Kansas high school mentioned at the beginning of this article, and it's our recommendation for other school sites and systems engaging in similar work.

Thus, our answer to "How do we get started with grading reform?" and "Our teachers completed a book study—what's the next step?": Start with the right *people*, establish a *purpose*, determine key *policies*, be mindful of *pacing*, and integrate support to sustain *persistence*.

Acknowledgments: The authors would like to thank Chad Lang for providing his helpful feedback on this manuscript.

REFERENCES

Brookhart, S. M. (2009). *Grading* (2nd ed.). Merrill.

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership, 69*(3), 10–14.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD.

Brookhart. S. M. (2017). *How to use grading to improve student learning*. ASCD.

Brookhart. S. M. (2026). Changes in K-12 grading research since 2016. *Journal of School Administration Research & Development, 11*(S1), 54-62.

Brookhart, S. M., Guskey, T.R., Bowers, A.J., McMillan, J.H., Smith, J.K., Smith, L.F., Stevens, M.T., & Welsh, M.E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. https://doi.org/10.3102/0034654316672069

Brookhart, S.M., & Guskey, T. R. (2019). Reliability in grading and grading scales. In T.R. Guskey and S.M. Brookhart (Eds.), *What we know about grading: What works, what doesn't, and what's next* (pp. 13-31).

Brookhart, S.M., & Chism, D. (2025). Equitable and culturally relevant grading practices. In *Culturally responsive assessment in classrooms and large-scale contexts* (1st ed., pp. 306–323). Routledge. https://doi.org/10.4324/9781003392217-21

Buckmiller, T., Townsley, M., & Cooper, R. (2020). Rural high school principals and the challenge of standards-based grading. *Theory and Practice in Rural Education, 10*(1), 92-102. https://doi.org/10.3776/tpre.2020.v10n1p92-102

Collins, J. (2001). *Good to great: Why some companies make the leap—and others don't*. HarperBusiness.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

DuFour, R., DuFour, R., Eaker, R., & Many, T. (2010). *Learning by doing: A handbook for professional learning communities at work* (2nd ed.). Solution Tree Press.

Griffin, R., & Townsley, M. (2022). Including homework and employability skills in class grades: An investigation of equitable outcomes in an urban high school. *Practical Assessment, Research, and Evaluation, 27*(1). https://doi.org/10.7275/pare.1327

Guskey, T. R. (2014, October 17). Why the label 'exceeds standard' doesn't work. *Education Week*. https://www.edweek.org/education/opinion-why-the-label-exceeds-standard-doesnt-work/2014/10

Guskey, T. R. (2020a). Breaking up the grade. *Educational Leadership, 78*(1), 40-46.

Guskey, T. R. (2020b). *Get set, go! Creating successful grading and reporting systems*. Solution Tree.

Guskey, T. R. (2023). Giving retakes their best chance to improve learning. *Educational Leadership, 80*(9). Online only. https://www.ascd.org/el/articles/giving-retakes-their-best-chance-to-improve-learning

Guskey, T. R. (2024). Addressing inconsistencies in grading practices. *Phi Delta Kappan, 105*(8), 52-57. https://doi.org/10.1177/00317217241251883

Guskey, T. R., Frey, N., & Fisher, D. (2024). *Grading with integrity: A research-based approach grounded in honesty, transparency, accuracy, and equity*. Corwin.

Hall, B. (2021). *Powerful guiding coalitions: How to build and sustain the leadership team in your PLC at work*. Solution Tree Press.

Hall, G.E., & Hord, S.M. (2020). *Implementing change: Patterns, principles, and potholes* (4th ed.). Pearson.

Harris, A., Jones, M., & Ismail, N. (2022). Distributed leadership: taking a retrospective and contemporary view of the evidence base. *School Leadership & Management, 42*(5), 438–456. https://doi.org/10.1080/13632434.2022.2109620

Heritage, M. (2022). *Formative assessment: Making it happen in the classroom* (2nd ed.). Corwin.

Kotter, J.P. (2012). *Leading change*. Harvard Business Review Press.

Kunnath, J., Townsley, M., & Radunzel, J. (2026). Silver lining or missed opportunity? A mixed methods study of COVID-19 school closures and grading reform efforts in K-12 schools. *Journal of School Leadership*. Online First. https://doi.org/10.1177/10526846261428221

Link, L. J., & Guskey, T. R. (2022). Is standards-based grading effective? *Theory into Practice, 61*(4), 406–417. https://doi.org/10.1080/00405841.2022.2107338

McLure, F. I., & Aldridge, J.M. (2023). Sustaining reform implementation: A systematic literature review. *School Leadership & Management, 43*(1), 70-98. https://doi.org/10.1080/13632434.2023.2171012

Morris, S. R., Parra-Martinez, A., Wal, J., & Maranto, R. (2026). The notorious SBG: Administrators' perceptions of standards-based grading practices. *Journal of School Leadership, 36*(1), 160-184. https://doi.org/10.1177/10526846251382329

O'Connor, K. (2022). *A Repair kit for grading: 15 fixes for broken grades.* FIRST Educational Resources, LLC.

O'Connor, K., Jung, L. A., & Reeves, D. (2018). Gearing up for FAST grading and reporting. *Phi Delta Kappan, 99*(8), 67-71. https://doi.org/10.1177/0031721718775683

Peters, R., & Buckmiller, T. (2014). Our grades were broken: Overcoming barriers and challenges to implementing standards-based grading. *Journal of Educational Leadership in Action, 2*(2). https://doi.org/10.62608/2164-1102.1070

Pfeffer, J., & Sutton, R. I. (2000). *The knowing-doing gap: How smart companies turn knowledge into action.* Harvard Business School Press.

Reeves, D. (2016). *FAST grading: A guide to implementing best practices*. Solution Tree.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.

Schimmer, T. (2016). *Grading from the inside out: Bringing accuracy to student assessment through a standards-based mindset.* Solution Tree.

Schimmer, T. (2023). *Redefining student accountability: A proactive approach to teaching behavior outside the gradebook*. Solution Tree.

Sinek, S. (2009). *Start with why: How great leaders inspire everyone to take action*. Portfolio/Penguin.

Spillane, J. P. (2006). *Distributed leadership*. Jossey-Bass.

Townsley, M. (2019). Considering standards-based grading: Challenges for secondary school leaders. *Journal of School Administration Research and Development, 4*(1), 35-38. https://doi.org/10.32674/jsard.v4i1.1941

Townsley, M. (2022). *Using grading to support student learning*. Routledge.

Townsley, M., Buckmiller, T., & Cooper, R. (2019). Anticipating a second wave of standards-based grading implementation and understanding the potential barriers: Perceptions of high school principals. *NASSP Bulletin, 103*(4), 281-299. https://doi.org/10.1177/0192636519882084

Townsley, M. & Knight, M. (2020). Making change stick: A case study of one high school's journey towards standards-based grading. *Journal of Educational Leadership in Action, 6*(3). https://doi.org/10.62608/2164-1102.1014

Townsley, M., & Kunnath, J. (2022). Exploring state department of education grading guidance during COVID-19: A model for future emergency remote learning. *Education Policy Analysis Archives, 30*,(163). https://doi.org/10.14507/epaa.30.7448

Townsley, M., & Sult, M. (2026). K-12 leaders' intentions to implement standards-based grading following professional learning. *Journal of Educational Leadership and Policy Studies, 10*. https://go.southernct.edu/jelps/files/2026-spring/12p_K_12_Leaders_Intentions.pdf

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*(3), 76-86. https://doi.org/10.1080/00098655.2024.2357615

Vatterott, C. (2015). *Rethinking grading: Meaningful assessment for standards-based learning*. ASCD.

Wilcox, J., & Townsley, M. (2022). Debunking myths of standards-based grading: Addressing the concerns and providing some strategies for implementing alternative grading practices. *The Science Teacher, 90*(1), 29-33.

MATT TOWNSLEY, EdD, is an Associate Professor of Educational Leadership in the Department of Learning, Leadership and Community at the University of Northern Iowa. His current research interests include educational leadership and classroom grading practices. Email: matt.townsley@uni.edu

JOSHUA KUNNATH, EdD, is an English teacher and instructional coach at Highland High School in Bakersfield, California. He also serves as editor-in-chief of the *Journal of School Administration Research and Development*. His current research interest is in grading reform. Email: joshkunnath@gmail.com

APPENDIX

RECOMMENDED READING: ESTABLISHING A COMMON GRADING REFORM KNOWLEDGE BASE

Topic	Resource
People	• Guskey, T. R. (2020). *Get set, go! Creating successful grading and reporting systems*. Solution Tree. • Hall, B. (2021). *Powerful guiding coalitions: How to build and sustain the leadership team in your PLC at work*. Solution Tree Press. • Kotter, J.P. (2012). *Leading change*. Harvard Business Review Press.
Purpose	• Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership, 69*(3), 10–14. • Guskey, T. R. (2020). *Get set, go! Creating successful grading and reporting systems*. Solution Tree. • Sinek, S. (2009). *Start with why: How great leaders inspire everyone to take action.* Portfolio/Penguin.
Policies	• Guskey, T. R., Frey, N., & Fisher, D. (2024). *Grading with integrity: A research-based approach grounded in honesty, transparency, accuracy,*

and equity. Corwin.
- O'Connor, K. (2022). *A Repair kit for grading: 15 fixes for broken grades*. FIRST Educational Resources, LLC.
- Reeves, D., Jung, L. A., & O'Connor, K. (2017). What's worth fighting against in grading? *Educational Leadership, 74*(8), 42-45.
- Schimmer, T. (2023). *Redefining student accountability: A proactive approach to teaching behavior outside the gradebook*. Solution Tree.
- Townsley, M. (2022). *Using grading to support student learning*. Routledge.
- 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*(3), 76-86. https://doi.org/10.1080/00098655.2024.2357615

***Special Issue**: Exploring Current & Future Grading Practices in K-12 Schools*
Research Article

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 74-83
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

Disrupting the Hidden Costs of Behavioral Grading: Uncovering Stakeholder Perceptions to Advance Equity

Laura J. Link
University of North Dakota, USA
Richard M. Simonds
University of Houston-Downtown, USA
Lizette A. Navarrete-Burks
University of Northern Colorado, USA

ABSTRACT

Grading practices in K-12 education continue to perpetuate inequities by including behavioral components that mask academic achievement and disproportionately affect historically marginalized students. This analysis examines perceptions of behavioral grading among five key stakeholder groups: administrators, instructional support staff, parents, teachers, and students. Comprehensive survey data were collected from over 20,000 participants in 10 U.S. school districts. Using GradingRx®'s validated assessment instrument, the study revealed significant perception gaps about the purposes of grading and whether behavioral factors should be included in academic evaluations. Teachers view grading as a tool for feedback and motivation. In contrast, students perceive grades as measures of inherent ability and gatekeepers to opportunities. The data reveal troubling disparities in stakeholder alignment, with administrators showing less agreement than other groups regarding behavioral grading practices. These findings underscore the urgent need for culturally responsive professional learning, transparent communication systems, and systematic reform of grading policies focused on equity and student voice. District and school leaders can use these stakeholder insights to transform grading from an administrative function into a strategic lever for educational justice and systemic change.

***Keywords*:** behavioral grading, educational equity, stakeholder perceptions, grading reform, cultural responsiveness

As schools across the United States confront persistent achievement gaps and mounting demands for educational equity, classroom grading practices have emerged as a critical focal point requiring immediate attention and systematic reform (Brookhart et al., 2016). The intersection of behavioral assessments with academic grading presents a particularly complex challenge that demands a nuanced understanding of stakeholder perspectives and their implications for student outcomes. Through our comprehensive work with GradingRx®, a leading provider of K-12 classroom assessment and grading survey instruments, we have uncovered significant disparities in how key educational stakeholders perceive the purposes and appropriateness of incorporating behavioral factors into academic grades (Duchesneau, 2020; Link et al., 2025). Two of the three authors work as consultants with GradingRx®, an educational consulting company that partners with school districts nationwide to assess and reform grading practices. This affiliation provides us with unique access to large-scale stakeholder perception data while also positioning us as advocates for grading reform guided by equity principles.

The practice of behavioral grading, incorporating elements such as participation, compliance, and classroom conduct into academic evaluations, represents more than a mere pedagogical choice; it embodies deeply embedded cultural assumptions that can either support or undermine educational equity (Rasooli et al., 2019). Research consistently demonstrates that behavioral grading practices frequently mask accurate academic achievement while disproportionately

impacting students from historically marginalized backgrounds (Backes et al., 2023; Di Liberto et al., 2022) This occurs when students receive lower grades not because of their mastery of content, but rather their perceived ability to conform to dominant cultural norms that may not align with their home experiences or cultural expressions (Wu et al., 2021).

Understanding stakeholder perceptions of behavioral grading becomes essential for school leadership seeking to implement equitable assessment practices. When perception gaps exist between those designing grading policies and those experiencing them, the potential for meaningful reform diminishes significantly (Link et al., 2025). This analysis examines comprehensive survey data from over 20,000 educational stakeholders across multiple U.S. districts to illuminate these perception disparities and their implications for advancing educational equity through grading reform.

This analysis addresses the following guiding questions that emerged from our ongoing work with district grading reform initiatives: First, to what extent do different stakeholder groups share common understandings about the primary communicative purposes of grades, particularly regarding information shared with parents? Second, how do stakeholders differ in their support for including behavioral factors (specifically behavior, attitude, and class participation) in academic grade determination? Third, what do these perception gaps reveal about the challenges and opportunities facing educational leaders who seek to implement more equitable grading practices? By examining these questions through analysis of comprehensive survey data, we aim to provide educational leaders with evidence-based insights that can inform culturally responsive approaches to grading reform.

The Cultural Context of Behavioral Assessment

The evaluation of student participation and behavior in academic settings is closely linked to broader cultural contexts. These contexts shape educational expectations and norms. Research shows that classroom participation expectations often reflect white, middle-class cultural values. These include verbal assertiveness, sustained eye contact, and engagement in teacher-directed activities (Delpit, 2006; Ladson-Billings, 2009; Yesbeck, 2011). When participation becomes a grading criterion, it may unintentionally favor students who display these dominant cultural behaviors. It can also disadvantage those from diverse backgrounds who show engagement through different cultural modalities (Paris & Alim, 2017).

Similarly, behavioral expectations embedded in grading practices frequently reflect implicit norms rooted in specific cultural values, such as deference to authority, limited physical movement, and restrained emotional expression (Skiba et al., 2011). These norms, while rarely articulated explicitly, serve as the foundation for behavioral expectations in many educational settings. When educators lack training to recognize and appreciate cultural variations in expression and engagement, student behavior is often interpreted through a deficit lens that reinforces existing inequities rather than acknowledging the validity of diverse cultural approaches to learning and participation (Gregory et al., 2010; Link & Guskey, 2019a).

The implications of these culturally embedded assessment practices extend beyond individual student outcomes to influence broader educational patterns. Students who cannot or choose not to conform to dominant cultural norms in their behavioral expressions may find themselves systematically disadvantaged in academic evaluations, regardless of their mastery of content or intellectual capabilities. This dynamic creates what we term “hidden costs” of behavioral grading—consequences that are not immediately apparent but accumulate over time to create significant barriers to educational opportunity and achievement.

Stakeholder Perspectives on Grading Purposes

Our comprehensive analysis of stakeholder perspectives reveals fundamental differences in how various groups conceptualize the primary purposes of grading. Teachers consistently view grading as a tool for providing feedback and motivating student effort, approaching assessment through a lens of instructional improvement and student development (Link, 2018; Link & Guskey, 2019b; McMillan, 2019). While teachers may intend for their grading practices to serve these formative purposes, research suggests that using grades as primary feedback mechanisms often misaligns with formative assessment principles, which emphasize feedback separated from evaluative judgment (Brookhart et al., 2016). This disconnect between teacher intentions and actual practice outcomes becomes particularly problematic when behavioral factors influence academic grades, further obscuring the feedback that might support student learning.

In contrast to teachers, students often perceive grades more as measures of their inherent abilities and as gatekeepers for future educational and career opportunities (Link & Hunter, 2024). This fundamental difference in perception creates a disconnect between educator intentions and student experiences that can undermine the motivational and learning benefits that grading is intended to provide. When students view grades as fixed judgments about their capabilities rather than feedback about their current performance, the potential for grades to support learning diminishes (Lipnevich & Smith, 2009).

Parents tend to view grades as indicators of both student achievement and school effectiveness, often regarding academic evaluations as measures of their child's progress and the quality of educational services provided (Link et al., 2025). This perspective adds another layer of complexity to grading reform efforts, as parental expectations and interpretations of grades may not align with either educator intentions or student experiences.

Instructional support staff, who often work closely with diverse student populations, may be more attuned to the potential inequities embedded in common grading practices (Da Fonte & Capizzi, 2015). Their proximity to students receiving additional academic support provides unique insights into how grading practices can either facilitate or hinder student success, particularly for those already facing educational challenges.

These divergent perspectives on grading purposes directly connect to stakeholders' willingness to include behavioral factors in academic evaluations. When stakeholders view grades primarily as communication tools for reporting student status to external audiences (parents, future educators, employers), the inclusion of behavioral information may seem appropriate and valuable. Conversely, when grades are conceptualized as feedback mechanisms supporting ongoing learning, behavioral factors that obscure academic achievement become problematic inclusions. The alignment, or misalignment, of stakeholder perspectives on both the communicative purposes of grades and the appropriateness of behavioral components, therefore, represents interconnected dimensions of the same fundamental challenge: establishing a shared understanding of what grades are meant to accomplish and what evidence legitimately belongs in determining them.

METHODS

Our comprehensive analysis employed convenience sampling (Shields et al., 2025) to examine survey data from 20,152 educational stakeholders across 10 geographically diverse U.S. school districts, collected through GradingRx®'s validated assessment instrument from March 2021 to August 2023. This extensive secondary dataset offers unprecedented insight into stakeholder alignment, or lack thereof, regarding fundamental grading practices and the integration of behavioral assessment.

Participant Demographics, Instrument, & Ethics

The stakeholder composition (Table 1) reflects a typical educational ecosystem, with students comprising the largest group (n = 11,605), followed by parents (n = 4,371), teachers (n = 3,571), instructional support staff (n = 340), and administrators (n = 265). Participating districts varied considerably in size and scope, serving between 1,800 and 23,000 students annually, with student participants representing grades 6-12 across both middle and high school levels.

Table 1

Role Characteristics of Stakeholders (N = 20,152)

Stakeholder Role	*n*	*%*
Administrators	265	1.3%
Instructional Support	340	1.7%
Parent	4,371	21.7%
Teacher	3,571	17.7%
Students	11,605	57.6%

Survey data collection utilized GradingRx®'s validated assessment instrument, which was developed through iterative piloting and refinement in partnership with educational researchers and practitioners. Survey items were drawn from scales developed in the research of Liu et al. (2006) and Guskey (2013). Pilot testing of the student, parent/guardian, and teacher/administrator surveys with samples of 20+ individuals each yielded internal reliability coefficients (α) of .81, .81, and .87, respectively, demonstrating that the items effectively captured meaningful variation in stakeholder perspectives on grading practices. The instrument employs four-point Likert scales (Strongly Disagree = 1 to Strongly Agree = 4), eliminating neutral positioning and requiring directional stakeholder perspectives on grading practices. Survey items were also drawn from established grading research frameworks (Brookhart, 2011; McMillan, 2001), with content validity established through expert review and pilot testing across diverse district contexts (GradingRx, 2025).

For purposes of this analysis, three survey items (questions 23, 33, and 37) addressing behavioral grading components were selected from the broader 38-item instrument. These items focused solely on parents' perceptions of the purpose of

grading and on two types of behavioral evidence used in determining grades (Appendix A). District-level and school-level administrators were combined into a unified category to facilitate meaningful cross-group comparisons.

This analysis utilized de-identified, aggregated survey data with no individual participant identifiers or personally identifiable information. The researchers secured proper permissions to conduct the survey research from a university's Institutional Review Board. All survey participants received informed consent documentation at the time of original data collection explaining that their aggregated, de-identified responses might be used for research and reporting purposes. The completely anonymous and aggregated nature of the data ensured that no individual participants could be identified in any analyses or reports.

Analytical Framework

Our analysis focused on three critical dimensions of behavioral grading practices, examining stakeholder perspectives captured through three specific survey items. Table 2 presents responses to survey item 23, which asked whether grades should primarily communicate information to parents about students' performance in school. Table 3 presents responses to survey item 37, which asked whether behavior or attitude should be used in determining students' grades. Table 4 presents responses to survey item 33, which asked whether class participation should be used in determining students' grades. The exact wording of these items, along with the full set of survey questions about grading purposes and evidence sources, appears in Appendix A. For each item, descriptive statistics including means, medians, standard deviations, and 95% confidence intervals were calculated and presented.

RESULTS & DISCUSSION

The data reveal concerning patterns in stakeholder alignment regarding the communicative purposes of grading and the inclusion of behavioral factors in academic evaluations. Across three key dimensions of behavioral grading, significant disparities exist, revealing fundamental challenges for equitable assessment and grading reform.

Communication to Parents and Grading Purposes

Students consistently showed the lowest within-group agreement ($M = 3.17$, $SD = 0.78$) that grading is meant primarily as communication to parents. This indicates a fundamental misalignment between how adults in the system conceptualize grades and those most directly impacted by grading practices (Table 2). In contrast, administrators ($M = 3.47$), instructional support staff ($M = 3.53$), parents ($M = 3.53$), and teachers ($M = 3.50$) demonstrate relatively high and consistent agreement with this communicative purpose. This disparity suggests that current grading practices may not effectively serve students' learning and motivational needs, despite being designed with these purposes in mind.

Table 2

Average Scores of Stakeholders' Beliefs Regarding Whether Grades Should Primarily Communicate Information to Parents About Students' Performance in School *(N = 20,152)*

Stakeholder	*Med*	*M*	*SD*	*CI* (95%)
Administrators	4	3.47	0.60	±.07
Instructional Support	4	3.53	0.57	±.06
Parents	4	3.53	0.62	±.02
Teachers	4	3.5	0.56	±.02
Students	3	3.17	0.78	±.01

Behavior and Attitude in Academic Assessment

The inclusion of behavior and attitude in grade determination reveals the most pronounced stakeholder divisions (Table 3). Administrators demonstrate notably lower agreement ($M = 2.38$, $SD = 1.14$) compared to all other stakeholder groups, with a difference of 0.94 scale points separating administrators from parents, which is nearly a full point on the four-point scale. This substantial gap in perspective suggests potential policy-practice disconnects that could fundamentally undermine reform efforts. The administrator group's mean falls closest to the midpoint between disagree and agree, while other

stakeholder groups' means cluster near agreement, indicating that administrators hold considerably more ambivalent or negative views.

Parents show the highest agreement (M = 3.32, SD = 0.82), followed by students (M = 3.21, SD = 0.86), instructional support staff (M = 3.01, SD = 0.96), and teachers (M = 2.91, SD = 0.98). This pattern suggests that while administrative leadership may acknowledge the equity concerns associated with behavioral grading, those closest to the classroom experience—parents, students, and support staff—tend to maintain stronger support for traditional practices that integrate behavioral assessments with academic evaluations.

Table 3

Average Scores of Stakeholders' Beliefs Regarding Whether Behavior or Attitude Should be Used in Determining Students' Grades *(N = 20,152)*

Stakeholder	*Med*	*M*	*SD*	*CI* (95%)
Administrators	3	2.38	1.14	±.14
Instructional Support	3	3.01	0.96	±.10
Parents	3	3.32	0.82	±.02
Teachers	3	2.91	0.98	±.03
Students	3	3.21	0.86	±.02

Examination of response distributions reveals that while all stakeholder groups show substantial variability in perspectives on behavioral grading, administrators demonstrate the widest spread of viewpoints, as evidenced by their standard deviation of 1.14. Parents show the tightest clustering of responses around agreement, with responses concentrated in the "agree" and "strongly agree" categories. Teachers and instructional support staff show more moderate distributions, with notable proportions selecting both agreement and disagreement options, suggesting divided perspectives within these professional groups on whether behavioral factors should be included in academic grades.

Class Participation in Grading Practices

Similar patterns emerge regarding the inclusion of class participation in grades (Table 4), with administrators showing significantly lower agreement (M = 2.83, SD = 0.97) compared to other stakeholder groups. Parents demonstrate the highest agreement (M = 3.37, SD = 0.68), followed closely by instructional support staff (M = 3.32, SD = 0.73), teachers (M = 3.29, SD = 0.72), and students (M = 3.21, SD = 0.81). The high agreement among parents, teachers, and instructional support staff on including participation in grading raises serious equity concerns, as it signals a collective reliance on traditional grading norms that can perpetuate cultural bias and systematically disadvantage students whose engagement does not align with dominant cultural practices (Gregory et al., 2010; Paris & Alim, 2017).

Table 4

Average Scores of Stakeholders' Beliefs Regarding Whether Class Participation Should be Used in Determining Students' Grades *(N = 20,152)*

Stakeholder	*Med*	*M*	*SD*	*CI* (95%)
Administrators	3	2.83	0.97	±.12
Instructional Support	3	3.32	0.73	±.08
Parent	3	3.37	0.68	±.02
Teacher	3	3.29	0.72	±.02
Students	3	3.21	0.81	±.01

IMPLICATIONS FOR LEADERSHIP AND REFORM

The perception gaps revealed by our comprehensive stakeholder analysis present both challenges and opportunities for educational leaders committed to advancing equity through grading reform. Our focus on descriptive patterns rather than inferential statistics reflects the exploratory nature of this multistakeholder investigation, where understanding the magnitude and direction of perception differences across five distinct groups provides essential groundwork for future hypothesis-driven research. The practical significance of these differences emerges not from statistical testing but from their

consistency across dimensions and their implications for policy implementation. The consistent disconnect between administrative perspectives and those of other stakeholders across all three dimensions (parent communication, behavioral assessment, and participation grading) suggests that current policy development processes may not adequately incorporate the voices and experiences of those most directly impacted by grading practices.

The particularly striking disparity regarding behavior and attitude inclusion (Table 3) underscores the complexity of reform efforts. While administrators appear to recognize the equity implications of behavioral grading, the higher agreement among parents, students, teachers, and instructional support staff reveals deeply embedded cultural expectations about appropriate student behavior and its relationship to academic evaluation. These findings suggest that behavioral grading practices may be so normalized within educational culture that stakeholders closest to the classroom experience view them as natural and necessary components of academic assessment.

Understanding why stakeholders support behavioral grading despite its documented inequities requires examining the situated logic that makes these practices seem reasonable from particular vantage points. Teachers working daily with students who appear disengaged may view behavioral grading as a legitimate tool for encouraging participation and effort, particularly when they lack alternative strategies for motivating struggling learners. Parents seeking a holistic assessment of their children may interpret grades that include behavioral components as more complete pictures of student development rather than contaminated measures of academic achievement. Students who have experienced behavioral grading throughout their educational careers may have internalized these practices as simply how school works, limiting their ability to envision alternatives without explicit education about the equity implications of grading. Even instructional support staff, despite their proximity to students facing educational challenges, may support behavioral grading because they see behavior and academic success as interconnected in students' daily experiences.

These perspectives are not simply wrong or uninformed; rather, they reflect genuine tensions in educational practice that grading reform must address. The challenge for educational leaders lies not in dismissing stakeholder support for behavioral grading as ignorance, but in creating professional learning experiences that help all stakeholders understand how practices that seem reasonable from their individual positions can perpetuate systemic inequities when examined from broader justice frameworks. Reform efforts must acknowledge the valid concerns that lead stakeholders to support behavioral grading while providing concrete alternative approaches that address those concerns without compromising equity.

District and school leaders must recognize that effective grading reform requires more than policy changes (Khalifa et al., 2016; Link, 2019); it demands a comprehensive cultural transformation that addresses deeply embedded assumptions about student behavior, engagement, and academic performance (Griffin & Townsley, 2021; Townsley, 2024). This transformation cannot occur without sustained investment in culturally responsive professional development that goes beyond surface-level cultural awareness to provide educators with skills for recognizing implicit biases, understanding structural inequities, and adopting inclusive instructional practices (Nieto & Bode, 2018).

The results in this study underscore the critical importance of engaging all stakeholder groups in meaningful dialogue about grading purposes and practices. When teachers, parents, support staff, and even students maintain stronger agreement regarding behavioral grading while administrators advocate for its removal, the potential for sustainable reform diminishes significantly. These disparities suggest that current professional learning systems may not be adequately preparing all stakeholders to understand the equity implications of traditional assessment practices or to implement alternative approaches that maintain academic rigor while honoring diverse cultural expressions of learning and engagement.

Additionally, the significant difference between student perspectives and those of adult stakeholders regarding parent communication underscores the urgent need to center student voice in grading reform efforts. Educational policies and practices that fail to align with student experiences and perceptions are unlikely to achieve their intended outcomes, regardless of their theoretical soundness or adult support (Charteris & Thomas, 2017). The finding that students show relatively high agreement with including behavior and attitude in grades suggests that young people themselves may have internalized traditional grading norms, potentially limiting their ability to envision more equitable alternatives without explicit education about the hidden costs of behavioral assessment.

Limitations

The descriptive nature of our survey data, while illuminating stakeholder perception patterns, cannot establish causal relationships or demonstrate that addressing these perception gaps will lead to the transformative outcomes suggested by our title. Our findings document what stakeholders believe about behavioral grading, but they do not capture the complex implementation dynamics that arise when districts attempt reform based on stakeholder input. The cross-sectional

convenience sample of 10 districts limits generalizability, and districts voluntarily engaging with GradingRx® may already possess greater equity awareness than the broader population of school systems. The aggregated data also prevents examination of variations within stakeholder groups by demographic characteristics or school contexts that may influence perspectives.

Our focus on three survey items captures only a narrow dimension of beliefs surrounding assessment and does not explore stakeholders' reasoning or willingness to support specific reforms. While we document concerning perception gaps, we cannot determine whether closing these gaps represents a necessary or sufficient condition for achieving equitable grading practices. Our contribution lies in documenting perceptual patterns that pose challenges for reform rather than in demonstrating solutions. The "disruption" and "equity advancement" in our title represent aspirational goals that our descriptive findings can inform but not accomplish independently.

Future Directions & Recommendations

Addressing the hidden costs of behavioral grading requires comprehensive, multi-faceted approaches that acknowledge the complexity of stakeholder perspectives while maintaining focus on equity and student success. District and school leaders must prioritize the development of transparent communication systems that facilitate ongoing dialogue between stakeholder groups and create opportunities for shared understanding of grading purposes and practices.

Professional development initiatives must move beyond traditional training models to embrace culturally expansive approaches that help educators recognize how their own cultural backgrounds influence their perceptions of appropriate student behavior and engagement. This training should provide concrete strategies for assessing student learning and participation in ways that honor diverse cultural expressions while maintaining academic rigor and high expectations for all students.

The systematic incorporation of student voice into the development of grading policies represents another critical component of meaningful reform. Students bring unique perspectives about how grading practices impact their learning, motivation, and educational aspirations that cannot be captured through the input of adult stakeholders alone. Creating formal mechanisms for student participation in grading reform ensures that policies and practices align with the experiences and needs of those they are intended to serve.

Additionally, district and school leaders must develop comprehensive feedback systems that enable ongoing monitoring of grading practices and stakeholder satisfaction. Regular assessment of stakeholder perceptions, combined with analysis of student outcome data, provides essential information for refining and improving grading reforms over time.

CONCLUSION

The comprehensive stakeholder data illuminates complex dynamics surrounding behavioral grading practices and their implications for educational equity. Significant perception gaps exist among stakeholder groups across all dimensions of behavioral assessment, underscoring the multifaceted challenges facing leaders committed to grading reform (Link & Mitic, 2025).

Most concerning is that administrators consistently have lower agreement with behavioral grading practices compared to all other stakeholder groups, while parents, teachers, instructional support staff, and students maintain high support for traditional approaches that integrate behavioral assessments with academic evaluations. This pattern suggests administrative reform efforts may encounter significant resistance from those closest to classroom experiences, who may not recognize the hidden costs these practices impose on historically marginalized students.

However, these challenges present unprecedented opportunities for transformation. By leveraging comprehensive stakeholder insights, district and school leaders can develop more effective and equitable assessment approaches that honor diverse student experiences while maintaining focus on learning and achievement. The path forward requires sustained commitment to culturally responsive practices, meaningful stakeholder engagement, and systematic policy reform guided by equity principles (Gay, 2018).

Through our continued work nationwide, we remain committed to supporting educational leaders in this critical transformation. The stakes are clear: inequitable grading practices undermine educational promise for countless students, while equitable assessment approaches unlock opportunity for all learners.

Ultimately, disrupting the hidden costs of behavioral grading represents more than technical reform; it embodies a fundamental commitment to educational justice, recognizing diverse ways students demonstrate learning, engagement, and growth. This requires courage, persistence, and unwavering focus on creating educational systems serving all students with excellence and equity, even when challenging deeply held assessment and grading beliefs.

REFERENCES

Backes, B., Cowan, J., Goldhaber, D., & Theobald, R. (2023). *Teachers and school climate: Effects on student outcomes and academic disparities.* Center for Analysis of Longitudinal Data in Education Research (CALDER). https://files.eric.ed.gov/fulltext/ED625572.pdf

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*, 3-12. https://doi.org/10.1111/j.1745-3992.2010.00195.x

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. https://doi.org/10.3102/0034654316672069

Charteris, J., & Thomas, E. (2017). Uncovering 'unwelcome truths' through student voice: Teacher inquiry into agency and student assessment literacy. *Teaching Education, 28*(2), 162-177. https://doi.org/10.1080/10476210.2016.1229291

Da Fonte, M. A., & Capizzi, A. M. (2015). Working collaboratively with support staff for inclusive education. In *Working with teaching assistants and other support staff for inclusive education* (pp. 195-218). Emerald Group Publishing Limited. https://doi.org/10.1108/S1479-363620150000004010

Delpit, L. (2006). *Other people's children: Cultural conflict in the classroom*. The New Press.

Di Liberto, A., Casula, L., & Pau, S. (2022). Grading practices, gender bias and educational outcomes: Evidence from Italy. *Education Economics, 30*(5), 481-508. https://doi.org/10.1080/09645292.2021.2004999

Duchesneau, N. (2020). Social, emotional, and academic development through an equity lens. *Education Trust Review*, 1 – 48. https://eric.ed.gov/?id=ED607298

Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice* (3rd ed.). Teachers College Press.

GradingRx® (2025). *Grading and assessment survey*. https://gradingrx.com/

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher, 39*(1), 59-68. https://doi.org/10.3102/0013189X09357621

Griffin, R., & Townsley, M. (2021). Points, points, and more points: High school grade inflation and deflation when homework and employability scores are incorporated. *Journal of School Administration Research and Development, 6*(1), 1-11. https://doi.org/10.32674/jsard.v6i1.3460

Guskey, T. R. (2013). Beyond tradition: Teachers' views of crucial grading and reporting issues. *Journal of Educational Research and Policy Studies, 13*(1), 32-49. https://eric.ed.gov/?id=ED509342

Khalifa, M. A., Gooden, M. A., & Davis, J. E. (2016). Culturally responsive school leadership: A synthesis of the literature. *Review of Educational Research, 86*(4), 1272-1311. https://doi.org/10.3102/0034654316630383

Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children* (2nd ed.). Jossey-Bass/Wiley.

Link, L. J. (2018). Teachers' grading perceptions: How pre-service training makes a difference. *Journal of Research in Education, 27*(2), 62-91. https://files.eric.ed.gov/fulltext/EJ1168160.pdf

Link, L. J. (2019). Leadership in grading reform. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading* (pp. 157-194). ASCD. https://www.ascd.org/books/what-we-know-about-grading?variant=118062

Link, L. J., & Guskey, T. R. (2019a). How traditional grading contributes to student inequities...and how to fix it. *Curriculum in Context, 45*(1), 12-19. https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1052&context=edp_facpub

Link, L. J., & Guskey, T. R. (2019b). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice, 26*(3), 303-320. https://doi.org/10.1080/0969594X.2018.1555515

Link, L. J., Guskey, T. R., & Lingat, J. M. (2025). Variation in stakeholders' views on homework and grading: A study to improve policy & practice. *Improving Schools, 26*(2-3), 73-98. https://doi.org/10.1177/13654802251325382

Link, L. J., & Hunter, C. A. (2024, April). *Students' perceptions of classroom assessment and grading*. [Paper presentation]. American Educational Research Association, Philadelphia, PA, United States.

Link, L. J., & Mitic, R. R. (2025). Examining grading purpose and practices: A multivariate analysis of secondary stakeholder perceptions. *Assessment in Education: Principles, Policies & Practice, 32*(3), 336–354. https://doi.org/10.1080/0969594X.2025.2534813

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied, 15*(4), 319–333. https://doi.org/10.1037/a0017841

Liu, X., O'Connell, A. A., & McCoach, D. B. (2006, April). *The initial validation of teachers' perceptions of grading practices.* [Paper presentation]. American Educational Research Association Annual Meeting, San Francisco, CA, United States. https://digitalcommons.lib.uconn.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1011&context=nera_2008

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20-32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x

McMillan, J. H. (2019). Surveys of teachers' grading practices and perceptions. In T. R. Guskey & S. M. Brookhart (Eds.), *What we know about grading* (pp. 84 - 112). ASCD.

Nieto, S., & Bode, P. (2018). *Affirming diversity: The sociopolitical context of multicultural education* (7th ed.). Pearson.

Paris, D., & Alim, H. S. (Eds.). (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. Teachers College Press.

Rasooli, A., Zandi, H., & DeLuca, C. (2019). Measuring fairness and justice in the classroom: A systematic review of instruments' validity evidence. *School Psychology, 22*, 701–722. https://doi.org/10.1080/2372966X.2021.2000843

Shields, S., Schartner, A., Curtis, W., & Murphy, M. (2025). *Research and education: An introduction to methods, approaches and processes* (2nd ed.). Routledge.

Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M. K., May, S. L., & Tobin, T. J. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review, 40*(1), 85-107. https://doi.org/10.1080/02796015.2011.12087730

Townsley, M. (2024). Dousing the flames of grading reform. *School Administrator, 81*(11), 30-34. https://www.aasa.org/resources/resource/dousing-the-flames-of-grading-reform

Wu, Z., Spreckelsen, T. F., & Cohen, G. L. (2021). A meta-analysis of the effect of values affirmation on academic achievement. *Journal of Social Issues, 77*(3), 702-750. https://doi.org/10.1111/josi.12415

Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors* [Doctoral dissertation, Virginia Commonwealth University]. ProQuest Dissertations and Theses Global. https://eric.ed.gov/?id=ED539229

LAURA J. LINK, EdD, is an Associate Professor of Teaching & Leadership at the University of North Dakota in the Department of Teaching, Leadership and Professional Practice. Her research focuses on improving K-12 grading leadership, teachers' grading practices, and the communicative value of grades for all education stakeholders. Her recent work addresses legal grading disputes and work habits in grades. With 35 years of K-12 and higher education experience, she has served as an administrator and educator across multiple levels. Email: laura.link@und.edu

RICHARD M. SIMONDS, EdD, is an Assistant Professor of Social Work at the University of Houston-Downtown. His primary research focuses on inequality and financial capability within social work and education. Email: simondsr@uhd.edu

LIZETTE NAVARRETE-BURKS, EdD, is an Assistant Professor of Educational Leadership in the Department of Educational Leadership and Policy Studies at the University of Northern Colorado. Her research examines equity-focused leadership, socially just practices in K–12 settings, and expanding access to high-quality STEM education for culturally and linguistically diverse students. Email: lizette.navarrete-burks@unco.edu

Manuscript submitted: ***August 12, 2025***
Manuscript revised: ***October 26, 2025***
Accepted for publication: ***November 8, 2025***

APPENDIX

GRADING AND REPORTING SURVEY

Items 23 – 28 asked about the purpose of grades. Specifically:

"The purpose of grades is …"

- Q23 to communicate information to parents about students' performance in school.
- Q24 to provide information to students about their learning progress.
- Q25 to select or group students for educational programs (Honors and Advanced classes, etc.).
- Q26 to provide incentives for students to learn.
- Q27 to evaluate the effectiveness of school programs.
- Q28 to recognize students' effort and responsibility.

Items 29 – 38 asked respondents to consider the sources of evidence teachers use in determining students' grades. Specifically, each item was stated:

"_____ should be used to determine students' grades."

Q29 Major exams and unit tests
Q30 Major projects and presentations
Q31 Class quizzes
Q32 Homework completion
Q33 Class participation
Q34 Work habits and neatness
Q35 Effort put forth (work ethic)
Q36 Punctuality of assignments
Q37 Behavior or attitude
Q38 Extra credit

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 84-89
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

The Retrograde Pendulum: Reversing Course on Standards-based Grading

Jay C. Percell
Illinois State University, USA

ABSTRACT

This commentary details one large, suburban school district that, after a decade-long commitment to implementing standards-based grading (SBG), reversed course, abandoned the initiative, and returned to traditional grading. This article provides a postmortem of the ambitious reform effort, drawing upon previous research and recent district survey data from teachers and parents. Among the factors leading to the reform's reversal are that, despite a carefully planned, multi-year rollout, the initiative was unable to resolve persistent concerns about implementation inconsistency, student motivation, and the technology reporting platform. Internal pressures were exacerbated by external turmoil, including personnel turnover and the COVID-19 pandemic, and district leadership ultimately acquiesced to widespread faculty and staff dissatisfaction with reporting measures, grade inflation, and ineffective communication of student progress. This district's experience offers important lessons for school leaders, emphasizing that systemic change requires fidelity of implementation accompanied by cultural adoption, long-term resilience and adaptability measures, and a commitment to instructional integrity.

***Keywords*:** standards-based grading, grading reform, educational leadership, grading practices

For over a century, the A-F grading system has been the primary language of academic achievement, one deeply embedded in school culture, and one that informs the identities of students and teachers alike (Guskey & Brookhart, 2019). It is against this backdrop that modern grading reform efforts, such as standards-based grading (SBG), have sought to shift the focus from accumulating points to providing specific, equitable feedback on student learning (Mehra, 2024; O'Connor, 2018). While SBG adoption has become widespread, the path to sustainable implementation remains challenging.

This commentary provides a postmortem of one such effort in the "Diamond Ridge School District" (pseudonym)— a large, diverse district previously featured as a case study during its initial SBG rollout (Percell & Meyer, 2021). After investing 10 years of time, research, and resources, district leadership abandoned its SBG grading reform and returned to traditional grading, prompted by dissatisfaction from parents, students, and teachers. This story highlights a critical lesson: For systemic change to endure, a well-intentioned plan must be accompanied by a profound and authentic shift in school and local community culture. Without deep cultural adoption, even the most promising reforms can fail.

INITIAL IMPLEMENTATION: LAYING THE GROUNDWORK FOR POTENTIAL PROMISE

Diamond Ridge leadership's decision to implement SBG reform was made through a deliberate, thoughtful, and patient process. Aware of the shortcomings of traditional grading methods and knowing a wide-reaching reform effort would take multiple years to implement correctly, district administration began to lay the groundwork for a strong foundation of understanding, open communication, and community support. These efforts included conducting district-wide book studies, hosting several parent-teacher information sessions, and creating a teacher-administrator task force charged with analyzing district grading practices and identifying leveled learning targets. Informational and professional development offerings on SBG began in 2014, two years before any actual reform implementation in Diamond Ridge schools. This development was undertaken to open community discourse on issues related to traditional grading practices and to communicate information regarding anticipated upcoming changes (Percell & Meyer, 2021).

After these informational offerings, Diamond Ridge initiated the grading rollout in a bottom-up manner, beginning with one district-wide subject area, elementary mathematics, piloting SBG in 2016. Teachers were given the option to implement SBG into other subjects as well, but it was not required. Year two of the rollout expanded SBG to all elementary-level subjects, and year three expanded SBG to all four of the district's middle and junior high schools. By year five, it was anticipated that the two Diamond Ridge high schools would also begin adopting SBG. However, leadership realized it might take more time to implement across all departments at the secondary level.

Despite careful planning and attempts at transparent communication of the reform's rationale, the SBG rollout faced resistance at every step. The district partnered with educational researchers, including the author, who collected longitudinal data from teachers, administrators, and parents throughout the implementation. It was this group that studied the Diamond Ridge reform effort as a case study, identifying common frustrations with SBG and making recommendations for potential resolutions (Percell & Meyer, 2021). Among them, five primary issues arose: teacher buy-in; logistical challenges of SBG gradebooks; securing parental buy-in; anxieties related to college acceptance; and student motivation. Many district teachers initially adopted a "wait-and-see" approach to SBG and never seemed to fully embrace the reform and its underlying rationale. Furthermore, district reporting platforms were initially designed for traditional grading and were ill-suited for reporting SBG.

Additionally, parents and local community members expressed confusion about SBG and a lack of trust in a system that removed the well-known, traditional A-F letter grades. They expressed uncertainties and fears that SBG would demotivate students and lead them to minimize effort, fumble genuine learning opportunities, and ultimately hinder their college and career readiness. One article highlighted the community's resistance at the time, citing general confusion and perceptions that SBG failed to individualize learning (Denham, 2019). While the case study sought to provide meaningful insight and potential resolutions for SBG issues, the ultimate decision to reverse course demonstrates that these challenges proved more lasting than initially anticipated.

MATRICULATING REFORM AND EXTERNAL DISRUPTIONS

The SBG reform initiative in Diamond Ridge faced more than just internal resistance; it also encountered a barrage of simultaneous external disruptions. As the rollout was just beginning in schools, district leadership changed. The Diamond Ridge superintendent resigned in 2019, and in the months that followed, several other key leadership positions turned over, including assistant superintendent, director of secondary schools, and chief financial officer. Diamond Ridge filled most of these roles with individuals who maintained a desire to see the grading reform through and still believed SBG was the most effective method for reporting students' specific skills related to learning targets and content standards, especially given the district's investment in the initiative to that point.

In addition to personnel changes, Diamond Ridge also switched to a new student information system (SIS). In 2019, the contract with their previous SIS provider expired. While the new SIS was not selected solely because of the ongoing SBG initiative, it likely played some part in the decision to choose an SIS touted as more closely aligned with SBG reporting than the previous system. As with any new software, integration and familiarity with new processes take time, and every piece of district information was affected by this change, not solely the teachers' learning management system and how their grades were imported.

Furthermore, amid this multi-layered reorganization, Diamond Ridge—along with the rest of the world—was blindsided by the 2020 COVID-19 pandemic. Like so many other districts, the pandemic halted face-to-face learning environments across Diamond Ridge and necessitated quickly pivoting to online instruction at every level—a pivot that district faculty and staff were not necessarily equipped to make, and one that the district was not necessarily prepared to facilitate. Nonetheless, like countless schools and districts across the country, the Diamond Ridge administration, faculty, and staff threw themselves heroically into meeting these challenges and doing their best to support their students, colleagues, and communities during a time of global crisis.

On a national level, the forcible move to remote and hybrid learning also exposed the authenticity of traditional grading, and best practices relating to grading in a time of crisis were called into question (Blackwelder, 2020). Many districts adopted "do no harm" approaches to grading, focusing on grace and flexibility over punitive measures (Hatch, 2020). While philosophically aligned with the inherent equity goals of SBG, this chaotic period likely added to the confusion in Diamond Ridge, blurring the lines between intentional, standards-based assessment and emergency-driven leniency.

RETURN TO NORMAL: A TRADITIONAL RESURGENCE

The return to in-person schooling did not bring a return to normalcy in Diamond Ridge so much as it brought unprecedented levels of teacher and student burnout and anxiety, along with a collective feeling of uncertainty about the very purpose of school. In this emotionally turbulent time, the desire to continue the complex and seemingly confusing grading reform initiative had cooled considerably. The logistical demands of communicating and understanding students' progress through a complicated LMS became burdensome for teachers, parents, and students throughout the district. The call to return to more traditional forms of grading intensified, as the local community sought to regain a measure of familiarity amid an overwhelmingly complex time.

Diamond Ridge administration held several meetings with the NEA-affiliated teachers association, and the district agreed to revisit the grading initiative. As part of a community strategic planning survey, public opinion was sought from parents, teachers, and certified staff members on a range of topics they felt Diamond Ridge should prioritize in its strategic planning over the next five years. The district administration designed the survey, and results were posted in a publicly accessible online space; well over 2,600 individuals responded. Although parental reactions to questions about the clarity of district grades and SBG were relatively neutral, there was a distinct consensus when asked about interest in forming an SBG focus group—an overwhelming 77% responded "no."

DATA OF DISSENT: TEACHERS REPORT SBG FAILINGS

The survey instruments for faculty and certified staff sought more specific feedback about the practical implementation of SBG across the district, and the responses were also very consistent. Teachers and staff voiced widespread dissatisfaction with how the SBG initiative was working. The most common note of disapproval concerned grade inflation. They felt that SBG was leading to rampant grade inflation, that their grades did "not accurately reflect what a student knows," and that grades were "not accurately displaying student academic performance."

Similarly, teachers were frustrated with the district gradebook and the way SBGs were being reported. Teachers overwhelmingly felt that letter grades associated with the SBG reports calculated by the district gradebook were not accurate reflections of students' abilities and sometimes were not even accurate reflections of their own entered SBG scores. One teacher respondent explained that the gradebook "does not seem to have the ability to correctly calculate grades in a SBG format." The teacher went on to voice their concern that "it is destroying student motivation and generally not achieving the goals that were set forth with this method of reporting."

One quantitative survey question asked whether teachers felt the summative standards-based grade in the district gradebook accurately reflected students' achievement, to which 60% responded negatively. Furthermore, nearly half of all teachers (49%) felt that a traditional gradebook would allow for more specific feedback regarding students' strengths and weaknesses related to learning standards (see Figure 1).

Teachers remarked that the enduring confusion and misunderstanding about SBG among parents, students, and even some teachers was leading to decreased faculty morale and disengagement from the surrounding community. One teacher remarked, "Parents don't understand it." They further explained, "Especially parents who had negative experiences in public schools, themselves. If they don't understand their kids' grades, then they feel even more detached from the learning process, which hinders our equity goals."

Furthermore, not only did teachers largely admit that SBG practices were being implemented inconsistently among departments and/or individual teachers, but that the entire grading system was negatively impacting students' academic achievement. Teachers argued the system was counterproductive, as students had "little understanding of how their grade was calculated" and became less engaged. One noted that the gradebook, whose "sole purpose is to clearly communicate student achievement," was instead holding back that communication. The teacher went on to say, "The SBG instruction and assessments have made solid improvements to our classes, but the reporting has been broken, flawed, and actively hurting students since we started doing it."

To that end, teachers and staff largely expressed continued belief in standards-based instruction, the fourth-highest category in teachers' qualitative responses. One such response conveys this belief: "I think the ideas behind SBG are good—targets, reassessments, minimization of formative assessments, but I cannot state enough how much I dislike SBG reporting."

Some teachers even reiterated their belief in SBG; however, the inconsistency in how the initiative was applied and the confusion it engendered outweighed any instructional benefit. One teacher stated,

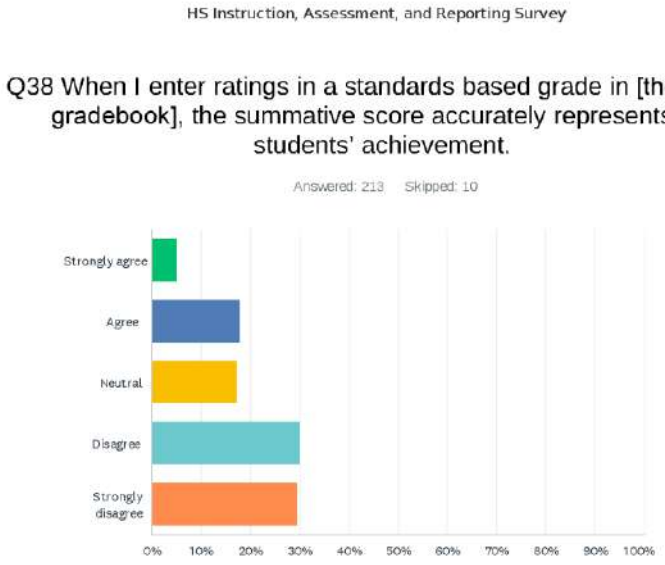
I wholeheartedly believe in standards-based teaching and standards-based assessment but cannot wrap my head around the ineffective attempt at standards-based reporting. The work we have done over the past 10 years to implement standards-based teaching and assessment has allowed us to focus on essential skills, making courses and assessment more consistent.

This teacher then continued,

We are doing everything we can to implement SBG in a way that is best serving our students. However, the reporting aspect is beyond our control. [The gradebook] is incredibly overwhelming for students and parents. They do not have a clear picture of how to improve or in what areas they are deficient—and that's what the rubrics are for.

Figure 1

High School Teacher Survey Responses on Grade Accuracy



ANSWER CHOICES	RESPONSES	
Strongly agree	5.16%	11
Agree	17.84%	38
Neutral	17.37%	37
Disagree	30.05%	64
Strongly disagree	29.58%	63
TOTAL		213

Finally, the teacher added, "I truly hope these surveys are being read and taken seriously. We have felt discouraged at the performative attempts seeking feedback that never actually result in trusting the teachers' expertise."

MESSAGE RECEIVED: THE DECISION TO REVERSE COURSE

The Diamond Ridge leadership was indeed listening, and they took the teacher and staff feedback seriously. An ad hoc grading task force was convened in the summer of 2024 to pore over the survey data and determine a path forward. The task force recommended ending the SBG initiative, and the district administration ultimately agreed, despite having invested a decade's worth of time and resources in fostering grading and assessment reform.

Beginning that fall, secondary departments were allowed to select their preferred grading format, either returning to a traditional gradebook or continuing with SBG. It was recommended that all teachers in the same department use the same grading format to maintain consistency, especially after the previous confusion and turmoil. In both high schools and all four junior high schools, every department throughout Diamond Ridge elected to return to traditional grading. As one teacher suggested, "Please consider giving teachers a choice of how they decide to report grades. I gave SBG reporting an

honest try over the last 8 years or so, and I honestly believe I was doing it better (using all the SBG tenets) prior to being forced to use SBG reporting."

IMPLICATIONS FOR DISTRICT LEADERSHIP

The story of Diamond Ridge is not necessarily a moratorium on standards-based grading. At the time of reversal, district leadership and most teachers still believed in the merits of standards-based instruction and standards-based assessment. Instead, this story speaks to the enormous challenge of enacting systemic change, especially during times of complex uncertainty, and it offers powerful lessons for other district leaders to consider.

Implication #1: Implementation Fidelity vs. Cultural Adoption

A plan can be executed perfectly, but if it fails to win the "hearts and minds" of the community, it cannot last. Such was the case with Diamond Ridge. The reform was well-intentioned and executed with a thoughtful, purposeful plan. However, it was never able to completely shift the prevailing culture among community members and even within the schools themselves. This impasse is emblematic of the Concerns-Based Adoption Model (CBAM), which asserts that individuals progress through predictable stages of concern when faced with a new initiative (Hall & Hord, 2020). District parents remained stuck in the highly personal stages of concern, focused on how SBG reform impacted their own children's GPAs and college readiness. Meanwhile, although district leadership provided information and rationale, they were unable to lead stakeholders to higher-level concerns, such as the reform's consequences for deeper student learning. This shows that communication must be a reflexive process that proactively identifies and addresses the specific concerns of each stakeholder group to help them move through deeper levels of adoption (Coburn, 2003).

Implication #2: Major Reforms Must be Built for Resilience, Not Rigidity

The already-stressed SBG initiative could not withstand the external shocks of a pandemic and the subsequent burnout crisis. It behooves district leaders to ensure that any complex reform efforts are designed with flexibility in mind. As Fullan (2007) deduced, sustainable reform efforts require adaptability. In a volatile educational landscape, initiatives need built-in "off-ramps" or points for modification that allow districts to scale back or adapt during periods of systemic stress without abandoning the project entirely. The Diamond Ridge vote to return to traditional grading was a vote for simplicity in a time of chaos. At the same time, a more resilient reform effort might have envisioned alternative variations or initiative modifications to curtail a complete course reversal.

Implication #3: Pedagogy Must Drive Technology, Not the Reverse

The widespread dissatisfaction with the district's mandated gradebook was a primary catalyst for the SBG reform's failure. This highlights a critical lesson: a reform is only as strong as the tools that support it. District leaders must ensure that technology effectively supports the instructional needs of their schools, and decisions about adopting technology tools must involve significant input from the teachers who will use them daily. Imposing a cumbersome or misaligned platform can create a constant source of friction that erodes teacher buy-in and turns pedagogical allies into frustrated adversaries.

CONCLUSION

The decision by Diamond Ridge School District to revert to traditional grading after a decade-long commitment to comprehensive SBG reform is a sober reminder of the gap that can exist between a reform's promise and its practical implementation. This district's experience underscores that successful, sustainable change is about more than just adopting a new model; it is a relentless process of communicating a vision, fostering a culture, managing resources and technology, and adapting to a constantly changing educational landscape. For other districts, the story of Diamond Ridge should not be a deterrent to pursuing equitable grading practices, but rather a roadmap to the very real and profound challenges that must be navigated to ensure such important work is sustainable, especially during times of uncertainty.

REFERENCES

Blackwelder, A. (2020, May 24). Grading in the new normal w/ Jay Percell. *Grow Beyond Grades*. https://growbeyondgrades.org/blog/episode-20

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12. https://doi.org/10.3102/0013189X032006003

Denham, R. (2019, December 11). New grading model met with confusion, pushback in unit 5 high schools. *WGLT.org*. https://www.wglt.org/show/wglts-sound-ideas/2019-12-11/new-grading-model-met-with-confusion-pushback-in-unit-5-high-schools

Fullan, M. (2007). *The new meaning of educational change (4th ed.)*. Teachers College Press.

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

Hall, G. E., & Hord, S. M. (2020). *Implementing change: Patterns, principles, and potholes* (5th ed.). Pearson.

Hatch, K. (2020, April 17). How do we grade during the shutdown? "Do no harm" is the answer. *Association of Washington Principals*. https://awsp.org/connect/blog/blog-detail/how-do-we-grade-during-the-shutdown-do-no-harm-is-the-answer/

Mehra, A. (2024, June 4). There's a more equitable way to grade: Districts should invest in it. *Edsurge*. https://edsource.org/2024/theres-a-more-equitable-way-to-grade-districts-should-invest-in-it/713086

O'Connor, K. (2018). *How to grade for learning: Linking grades to standards* (4th ed.). Corwin.

Percell, J. C., & Meyer, B. B. (2021). Resolutions for a new paradigm: Addressing common issues in standards-based grading. *Kappa Delta Pi Record*, *57*(4), 185–190. https://doi.org/10.1080/00228958.2021.1968692

JAY C. PERCELL, EdD, is an Associate Professor in the School of Teaching and Learning at Illinois State University. His major research interests are comprised of disciplinary and digital literacies, alternative grading, secondary teacher education, and AI Literacy in the secondary classroom. Email: jpercel@ilstu.edu

Manuscript submitted: ***September 30, 2025***
Manuscript revised: ***January 23, 2026***
Accepted for publication: ***January 26, 2026***

***Special Issue**: Exploring Current & Future Grading Practices in K-12 Schools*
Research Article

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 90-104
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

From Fragmented to Aligned: A Standards-Based Grading Continuum in Teacher Preparation

Keegan VanDevender
Drake University, USA

ABSTRACT

Standards-based grading (SBG) has gained traction across PK–12 education as a more transparent and mastery-oriented approach to assessing student learning. However, considerably less is known about how undergraduate teacher education programs (TEPs) prepare pre-service teachers to enter schools where SBG is practiced. This qualitative multiple-case study examined how three undergraduate TEPs in a Midwestern state integrated instruction on grading and assessment into their curricula. Drawing on interviews with teacher education faculty and analysis of programmatic documents, cross-case analysis revealed patterned variation in how grading instruction was embedded across coursework and field experiences. From these findings, this study inductively developed a Continuum of Grading Approach Integration, which illustrates three approaches to teacher preparation: fragmented passive exposure, emerging integration, and an aligned programmatic approach. The continuum serves as a heuristic for understanding how TEPs vary in their support of candidates' readiness to engage in SBG practices that are increasingly present in PK–12 schools. Implications are discussed for teacher education faculty, program leaders, and educational leadership seeking to align teacher preparation with evolving grading reforms.

***Keywords*:** standards-based grading, teacher preparation, teacher education, grading reform, qualitative case study, educational leadership

Across PK–12 education, grading practices have become a central focus of reform as educators, families, and policymakers question whether traditional percentage- and letter-based systems accurately communicate student learning (Brookhart, 2015; Guskey, 2013; Townsley & Buckmiller, 2020). Longstanding critiques have documented how conventional grading practices often conflate academic achievement with non-cognitive factors such as behavior, compliance, and effort, resulting in distorted representations of student mastery and inequitable outcomes (Link, 2019). In response, standards-based grading (SBG) has gained traction as an alternative approach that emphasizes transparency, criterion-referenced assessment, and communication of learning relative to clearly defined standards (Guskey & Jung, 2013; Schimmer et al., 2018; Townsley & Buckmiller, 2020).

As SBG continues to expand across PK–12 educational contexts, attention has increasingly turned to questions of implementation, coherence, and sustainability (Guskey, 2015; Townsley, 2020). Research has documented both the promise and complexity of grading reform, noting that shifts toward standards-based systems require changes not only in technical grading practices but also in professional beliefs, instructional coordination, and organizational structures (Brookhart et al., 2016; Guskey, 2001; Tierney et al., 2011; Townsley et al., 2019). While this literature has contributed important insights into how practicing teachers and school leaders navigate grading reform, considerably less attention has been paid to how teacher preparation programs equip pre-service teachers to interpret and enact grading practices prior to entering the profession (Battistone et al., 2019; Brookhart, 2017).

Teacher education programs (TEPs) play a critical role in shaping how novice teachers understand assessment, grading, and their purposes within teaching and learning (Brookhart, 2017; Guskey, 2015). Yet existing research suggests that

preparation related to grading and assessment is often fragmented, inconsistently emphasized, or left to practicum experiences rather than addressed coherently across coursework (Barnes, 1985; Battistone et al., 2019). As a result, pre-service teachers may enter PK–12 schools implementing SBG without a clear understanding of its underlying principles or how grading practices align with instructional decision-making (Tierney et al., 2011). This disconnect raises questions about how TEPs are positioned to support grading reforms unfolding in contemporary PK–12 settings (Guskey, 2015; Townsley, 2020).

To address this gap, this qualitative multiple-case study explored how undergraduate TEPs integrate instruction related to grading and assessment within their respective curricula in contexts where SBG is increasingly present in PK–12 schools (Battistone et al., 2019; Guskey & Jung, 2013). Through cross-case analysis of interview data and programmatic documents, this study examined how teacher preparation programs structure grading-related learning opportunities, identifying patterned variation in how instruction is emphasized, coordinated, and enacted across coursework and field experiences (Saldaña, 2016; Stake, 2006).

From these findings, the study inductively developed a Continuum of Grading Approach Integration that illustrates three approaches to teacher preparation: fragmented passive exposure, emerging integration, and aligned programmatic approach (Miles et al., 2020). The continuum is offered as a heuristic for understanding how TEPs may differentially support pre-service teachers' readiness to engage in SBG practices, rather than as an evaluative scale or prescriptive model. By centering the preparation pipeline, this study extends existing grading reform literature beyond PK–12 implementation to illuminate how TEPs may shape the future enactment of grading practices in schools.

Research Questions

Guided by this focus, this study applied the following research question: *How are undergraduate teacher education programs (TEPs) integrating instruction related to grading and assessment practices within their curricula?* This question centers on programmatic integration across coursework and field experiences, illuminating variation in how teacher preparation programs structure grading-related learning opportunities.

REVIEW OF LITERATURE

Grading has long occupied a contested position in educational systems, functioning simultaneously as a mechanism for measuring student learning, communicating achievement, and regulating behavior within a single, often ambiguous evaluative signal (Barnes, 1985; Brookhart, 2011; Guskey, 2015). Foundational scholarship has consistently critiqued traditional grading practices for conflating academic achievement with non-academic factors such as effort, compliance, and behavior, thereby undermining the validity and interpretability of grades as indicators of learning and weakening their usefulness for instructional decision-making (Guskey, 2001; Kohn, 2011). From a measurement perspective, such practices obscure what grades represent, limiting their instructional usefulness and complicating decision-making for students, families, and educators by masking what students actually know and are able to do (Brookhart, 2017). Despite these longstanding debates, relatively little attention has been given to how these tensions are addressed within teacher preparation programs.

At the same time, grades operate as institutional signals that carry high-stakes consequences related to promotion, placement, and postsecondary access despite their inconsistent meaning across contexts (Guskey, 2015). Tierney et al. (2011) described this dual role as a persistent source of tension within grading systems, noting that grades are often expected to serve multiple, and sometimes competing, purposes without clear articulation of priorities or shared agreement about what grades are intended to communicate. This ambiguity contributes to wide variation in grading practices across classrooms and schools and raises important questions about how pre-service teachers are prepared to navigate these inconsistencies (Brookhart, 2011; Guskey, 2015).

Standards-Based Grading and the Challenge of Coherent Implementation

SBG has emerged as a prominent reform-oriented response to longstanding critiques of traditional grading systems, yet most research has focused on its implementation in PK–12 settings rather than how educators are prepared to enact these practices (Guskey & Jung, 2013; Marsh, 2023). At its core, SBG emphasizes clearly articulated learning standards, criterion-referenced assessment, and grading practices intended to communicate student learning more accurately by separating academic achievement from non-cognitive factors (Guskey & Jung, 2013; Schimmer et al., 2018). Proponents have argued that by centering grades on evidence of learning rather than point accumulation, SBG can enhance transparency and support

instructional decision-making for both teachers and students (Brookhart, 2011; Guskey, 2001; Marsh, 2023). However, these expectations assume that teachers enter the profession with a clear understanding of how to implement such practices, an assumption that has received limited attention in teacher preparation research.

Despite its conceptual clarity, research has consistently demonstrated that SBG is difficult to implement cohesively across classrooms, schools, and systems, raising questions about how educators are prepared to interpret and enact grading practices prior to entering the profession (Guskey, 2015). Studies of school- and district-level reform described persistent variation in how grading principles revealed in policy are interpreted and enacted in practice, even when formal policies are in place (Townsley, 2020). Recent work reinforced this finding, emphasizing that implementation challenges often stem from limited shared understanding among educators rather than from technical features of grading systems themselves (Townsley & Wilcox, 2024; Zhang, 2024). Additionally, misunderstandings about the purposes and mechanics of SBG continue to shape inconsistent enactment, particularly when professional learning is fragmented or disconnected from daily instructional practice (Zhang, 2024).

Together, this body of scholarship suggested that grading reform is not merely a technical change in reporting practices but a broader organizational and cultural shift that requires consistency across instructional systems and sustained professional learning structures (Schimmer et al., 2018; Tierney et al., 2011). Without such coherence, reform efforts risk becoming superficial or unevenly applied, underscoring the importance of examining how teachers are prepared to engage with grading practices before entering PK–12 classrooms (Guskey, 2015).

Equity-Oriented Perspectives and Contemporary Debate

Equity has become an increasingly central lens for grading scholarship, with researchers arguing that traditional grading practices may reproduce inequities by privileging students who are more familiar with dominant cultural norms and school expectations embedded in behavioral compliance and hidden-curriculum structures (Brookhart, 2011; Feldman, 2019). From this perspective, equity-oriented grading emphasizes transparency, consistent criteria, and opportunities for students to demonstrate learning over time, aligning closely with standards-based approaches that prioritize mastery rather than sorting (Schimmer et al., 2018). These concerns further highlight the need to examine how teacher preparation programs introduce and frame equity-oriented grading practices for pre-service teachers.

However, both foundational and recent literature cautions against assuming that reform language alone produces equitable outcomes. Kohn (2011) critiqued grading reforms that retain hierarchical structures while rebranding practices as more equitable, arguing that such approaches may obscure deeper systemic issues related to competition, ranking, and compliance-driven schooling (Kohn, 2011). Empirical studies similarly highlight variation in teacher support for equity-oriented grading practices, suggesting beliefs about fairness, rigor, and accountability remain contested even among educators committed to reform principles (Morris et al., 2023).

Recent policy-oriented analyses echo this tension, documenting skepticism and concern among educators regarding the implementation and consequences of equitable grading initiatives, particularly when reforms are perceived as externally imposed (Fordham Institute, 2025; Link & Guskey, 2022; Morris et al., 2023). This body of research suggests that equity in grading is contingent on shared understanding, consistency across practice, and sustained professional learning across instructional systems. Guskey (2015) emphasizes that without consistent interpretation and application of grading principles, reform efforts may exacerbate confusion rather than mitigate inequities for both educators and students. This insight underscores the importance of examining how educators are prepared to engage with grading practices before they enter PK–12 classrooms. Taken together, this scholarship positions equity not as a fixed outcome of grading reform but as contingent on shared understanding and consistent practice, raising questions about how these concepts are introduced and developed within teacher preparation programs.

Teacher Preparation, Assessment Literacy, and the Preparation Gap

Despite extensive research on grading reform in PK–12 contexts, comparatively little scholarship has examined how teacher preparation programs explicitly address grading and assessment practices as a coherent component of professional preparation (Battistone et al., 2019; Brookhart, 2017). Brookhart (2017) noted that preservice education often emphasizes assessment broadly while devoting limited attention to grading decision-making, reporting practices, and the purposes grades are intended to serve, a pattern that mirrors longstanding ambiguity in the field regarding the instructional and communicative functions of grades (Guskey, 2015; Tierney et al., 2011). As a result, pre-service teachers may enter classrooms without a coherent framework for interpreting grading practices, particularly in contexts where SBG challenges

traditional assumptions about assessment, feedback, and accountability (Battistone et al., 2019; Schimmer et al., 2018; Townsley & Buckmiller, 2020).

Research on grading reform has suggested that teachers' understandings of grading are shaped through a combination of coursework, clinical experiences, and professional socialization across multiple stages of their preparation and induction into the profession (Battistone et al., 2019; Guskey et al., 2020; Townsley & Buckmiller, 2020). When preparation programs do not provide coherent frameworks for interpreting grading practices, preservice teachers may rely heavily on mentor teachers' practices during field placements, which vary widely in alignment with reform-oriented principles and often reflect localized norms rather than research-informed grading approaches (Brookhart et al., 2016; Tierney et al., 2011). Recent studies have reinforced this concern, indicating that without intentional support, preservice teachers often experience grading as situational and context-dependent rather than as an integrated component of instructional decision-making linked to standards, assessment design, and feedback practices (Lewis, 2022; Zhang, 2024).

Emerging research has also suggested that variation in preparation may shape how educators later respond to grading reform initiatives as they assume roles with increased responsibility for instructional and assessment leadership (Guskey, 2015; Link, 2019; Morris et al., 2025; Townsley, 2020). Morris et al. (2025) found that administrators' perceptions of SBG were influenced by prior experiences and preparation, highlighting the long-term implications of how grading is learned within professional pathways. Taken together, this literature positions teacher preparation as a critical, yet underexamined site for understanding how grading practices are interpreted, normalized, and enacted over time, particularly as grading reforms continue to evolve across PK–12 systems (Brookhart, 2017; Schimmer et al., 2018).

Toward a Programmatic Lens on Grading Preparation

Across the grading reform literature, coherence repeatedly emerges as a necessary condition for meaningful change and sustainable implementation (Guskey, 2015; Schimmer et al., 2018). While scholars have examined coherence at the classroom, school, and district levels, far less attention has been given to how coherence is cultivated within teacher preparation programs as organizational systems. Existing research suggests that fragmented or inconsistent preparation may shape how novice teachers make sense of grading practices, yet the structure and integration of grading instruction across preparation curricula remains largely unexplored in empirical research.

Addressing this gap requires a programmatic lens that examines how grading and assessment instruction is intentionally structured, coordinated, and reinforced across coursework and field experiences within teacher preparation programs. Despite its importance, limited research has examined how teacher preparation programs organize and coordinate grading instruction at the program level. By examining variation in how undergraduate TEPs approach grading preparation, this study extends grading reform scholarship upstream to the preparation pipeline. In doing so, it responds to longstanding critiques in the literature regarding coherence, equity, and implementation, while offering a framework for understanding how preparation contexts may shape the future enactment of grading practices in PK–12 schools through newly licensed teachers.

Conceptual Framework

This study is informed by key tenets associated with SBG, including an emphasis on clearly articulated learning standards, criterion-referenced assessment, the separation of academic achievement from non-cognitive factors, and opportunities for students to demonstrate mastery over time (Guskey et al., 2020; Tierney et al., 2011). Rather than focusing on isolated courses or individual instructional strategies, this perspective emphasizes how programs create coherence in grading-related learning experiences throughout teacher preparation. In this study, programmatic integration refers to the extent to which grading and assessment instruction are intentionally structured, coordinated, and reinforced across coursework, faculty practices, and field experiences within a teacher preparation program. In keeping with a constructivist qualitative approach (Lincoln & Guba, 1985), the conceptual framework provided an analytic orientation for interpreting variation across cases, while allowing patterns, themes, and relationships to emerge inductively from participant perspectives and programmatic documents.

METHODS

This study employed a qualitative multiple-case study design (Stake, 2006) to examine how undergraduate teacher education programs integrate instruction on grading and assessment into their curricula. The design supported cross-case comparison of programmatic approaches. Consistent with a constructivist epistemological stance (Crotty, 1998/2015), the study centered

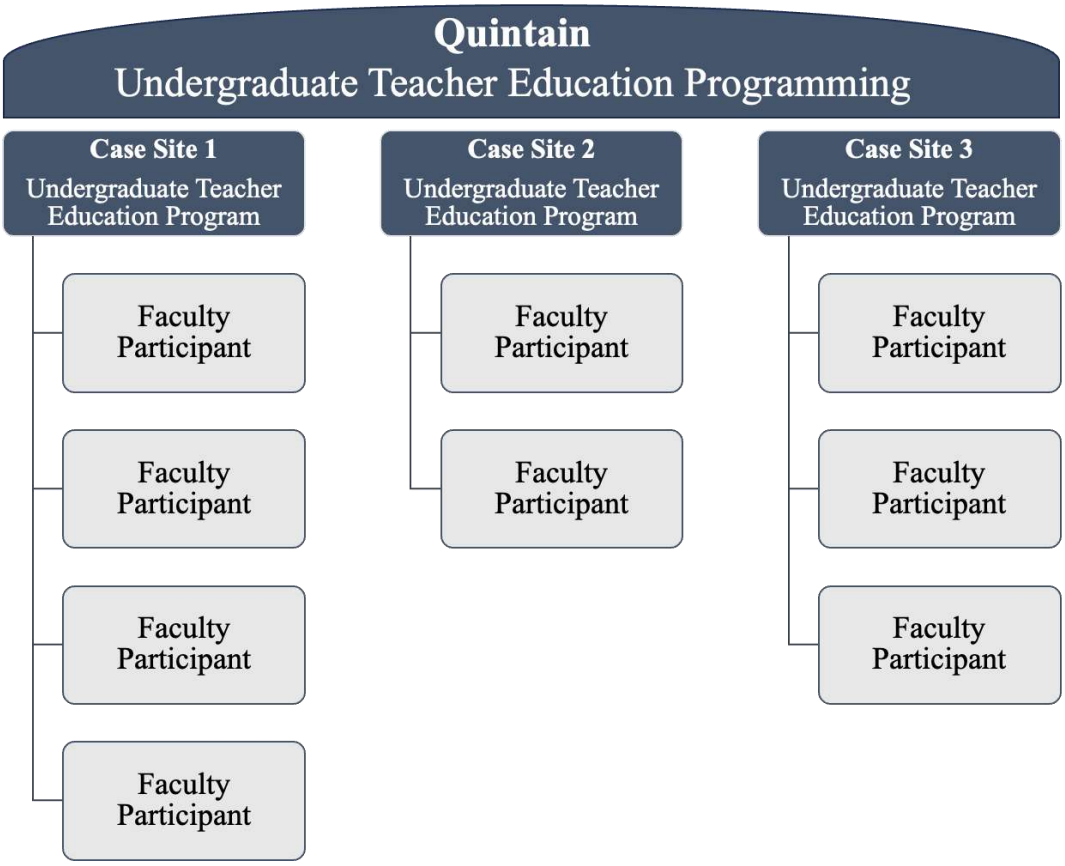
on participants' interpretations and programmatic contexts in understanding how grading-related instruction is organized and enacted.

The Quintain

The study was organized around a central quintain: how undergraduate teacher education programs prepare pre-service teachers for SBG environments. This concept guided case selection and cross-case analysis (Stake, 2006).

Figure 1

The Quintain: Undergraduate Teacher Education Programming



Data Collection

Data collection incorporated three primary sources: interview data, programmatic documents, and analytic memos (Merriam, 1988; Merriam & Tisdell, 2016; Stake, 2006; Yin, 2018). Using a modified version of Seidman's (2019) three-interview protocol, nine faculty participants each engaged in two semi-structured interviews. The first interview focused on faculty roles and experiences, while the second examined instructional practice, institutional context, and perceptions of grading (Seidman, 2019). Interviews were conducted via Zoom or in person, ranged from 45–90 minutes, and were audio-recorded and transcribed for analysis in NVivo (Version 14). Programmatic documents analyzed to triangulate findings included course syllabi, student teaching handbooks, and programmatic frameworks (Merriam, 1988; Stake, 2006; Yin, 2018), and analytic memoing was used to document reflections and support thematic development (Saldaña, 2016).

Trustworthiness, Ethics, and Positionality

Trustworthiness was supported through triangulation across data sources, member checking, and analytic memos documenting interpretations. An audit trail and peer debriefing were also used to support analytic rigor (Lincoln & Guba, 1985; Stake, 2006). All research activities were approved by the Institutional Review Board (IRB) at the researcher's home institution and by the site-based IRB at the individual case sites prior to data collection. Informed consent was obtained from all participants, who were advised of the study's purpose, their right to withdraw, and confidentiality protections. Pseudonyms were used, identifying details were removed, and job titles were generalized to protect anonymity. Data were securely stored on encrypted devices and in locked storage accessible only to the researcher.

The researcher brought a background in higher education and an interest in equitable grading practices within teacher preparation contexts. While this positionality informed an understanding of curriculum and faculty decision-making, it also introduced potential bias. To mitigate this, the researcher engaged in reflexive memoing throughout data collection and analysis and used peer debriefing to examine assumptions and interpretations. Transparency was further maintained by keeping a detailed research journal that provided an audit trail of the design decisions and personal reflections throughout the study. Through these methods, the study was purposefully designed to foreground participant perspectives and institutional contexts, ensuring the findings remained grounded in the data.

Participants and Case Sites

Case sites (see Table 1) and participants (see Table 2) were selected using purposeful sampling to capture variation in how SBG concepts are addressed within teacher preparation programs (Merriam & Tisdell, 2016). Sites were selected to reflect shared institutional characteristics (e.g., private, liberal arts, licensure-granting programs) while varying in program structure, faculty organization, and approaches to grading and assessment instruction. This approach supported cross-case comparison of how program design shapes candidates' opportunities to engage with grading practices.

Case Site Overview

The three institutions included in this study, referred to by the pseudonyms Cottonwood University, Sumac College, and Hackberry College, shared common features such as private liberal arts status, religious affiliation, and teacher licensure pathways, while also reflecting meaningful differences in faculty size, curricular structures, and institutional cultures (see Table 1). Institutional characteristics, including undergraduate enrollment and faculty composition, were drawn from publicly available data reported through the College Navigator database (National Center for Educational Statistics, 2022). Student-to-faculty ratios were calculated using full-time faculty counts only, as reported by the source, and do not include part-time faculty. This balance supported both within-case depth and cross-case comparison.

Table 1

Institutional Data by Case Site

Institution Pseudonym	Undergraduate Enrollment	Full-Time Faculty	Part-Time Faculty	Student-to-Faculty Ratio
Cottonwood University	1,404	47	79	18:1
Sumac College	1,153	77	123	15:1
Hackberry College	1,169	91	8	13:1

Participant Overview

The study included nine full-time faculty members across three small, private liberal arts colleges in a Midwestern state. Participant pseudonyms were self-selected, and job titles were generalized to ensure anonymity (see Table 2). The sample reflected variation in teaching experience, academic rank, and disciplinary specialization. Collectively, participants offered

insight into how individual beliefs, departmental norms, and institutional contexts shaped their approaches to grading and assessment within the preparation of future teachers.

Cottonwood University employed four participants representing diverse academic and professional backgrounds, including Social Studies education, literacy, and technology integration. Sumac College contributed two participants, both of whom held leadership roles in program coordination and assessment. Hackberry College included three participants with specializations in early childhood, mathematics education, and educational psychology.

Table 2

Faculty Participant Profiles by Institution

Cottonwood University Faculty Profile					
Pseudonym	Gender	Title	Highest Degree	Years in TEP	Years in PK–12
Arthur	Male	Associate Professor	Doctorate	10	5
Lynn	Female	Assistant Professor	Master's	8	16
Bob	Male	Assistant Professor	Master's	10	12
Anna	Female	Assistant Professor	Master's	18	11
Sumac College Faculty Profile					
Pseudonym	Gender	Title	Highest Degree	Years in TEP	Years in PK–12
Neice	Female	Assistant Professor	Doctorate	4	15
Lew	Male	Assistant Professor	Doctorate	9	16
Hackberry College Faculty Profile					
Pseudonym	Gender	Title	Highest Degree	Years in TEP	Years in PK–12
Emily	Female	Assistant Professor	Doctorate	4	15
Patricia	Female	Lecturer	Master's	2	6
Jane	Female	Associate Professor	Doctorate	11	5

Case Site 1: Cottonwood University

Cottonwood University is a Christian liberal arts college with a teacher education program housed within its School of Education and Human Services. The program offers a range of licensure areas and emphasizes clinical partnerships with local PK–12 districts. Cottonwood contributed four participants to the study. These faculty members represented various content areas, including literacy education, social studies education, educational technology, and secondary methods. Participants varied in academic rank and appointment status. Cottonwood's program demonstrated a strong emphasis on developmental coursework and included multiple practicum experiences prior to student teaching.

Case Site 2: Sumac College

Sumac College is a small Christian liberal arts college with a tight-knit education department and a cohort-based licensure pathway. The program is known for its emphasis on moral and civic education, which is integrated into coursework and field experiences. Two faculty members from Sumac participated in the study. Both held leadership roles within the program, one serving as a field experience coordinator and the other as an assessment coordinator, and both were involved in course design and program evaluation. The small faculty size at Sumac created conditions for alignment across courses, though formal grading frameworks were not explicitly embedded in all program documents.

Case Site 3: Hackberry College

Hackberry College is also a Christian liberal arts institution, yet it has a larger education department relative to its institutional size. Its teacher preparation program emphasizes early field experiences and differentiated instruction. Three faculty members from Hackberry participated in the study, each representing different areas of expertise: mathematics education, early childhood education, and educational psychology. Participants varied in academic rank and professional

experience, including prior PK–12 teaching and administrative roles. While faculty expressed support for equitable assessment practices, Hackberry demonstrated variability in how and where grading instruction appeared in the curriculum.

Data Analysis

Data analysis occurred concurrently with data collection and followed an iterative, inductive process consistent with qualitative multiple-case study methodology (Stake, 2006). Interview data (transcripts and notes) and programmatic documents were first reviewed within each case to develop a contextual understanding of how grading and assessment instruction was described and enacted within individual TEPs. Initial coding (Saldaña, 2016) was used to identify segments of data related to grading practices, assessment purposes, curricular emphasis, and connections across coursework and field experiences. These codes were then refined and grouped into broader categories within each case.

Comparisons across participants within institutions were used to examine consistency and variation in how grading-related instruction was positioned programmatically. Analytic memos supported the development of interpretations and identification of emerging themes (Stake, 2006). Cross-case analysis then compared these categories across the three institutional cases to identify patterns in grading preparation. Through constant comparison, three distinct approaches to programmatic integration emerged, reflecting variation in how grading and assessment instruction was coordinated across coursework, modeled by faculty, and reinforced through field experiences. The Continuum of Grading Approach Integration was developed as an analytic synthesis of these patterns, representing a data-grounded heuristic rather than a prescriptive framework.

Throughout the analysis, distinctions were maintained between descriptive findings grounded in participant data and interpretive claims developed through cross-case comparison. This approach enhanced transparency in how analytic claims were derived and allows the continuum to remain anchored in interview data and programmatic documents rather than imposed theoretical categories (Lincoln & Guba, 1985; Stake, 2006).

FINDINGS

Analysis across the three institutional cases revealed patterned variation in how grading and assessment instruction was integrated within undergraduate teacher education programs. While faculty across all cases acknowledged the growing prevalence of SBG in PK–12 contexts, they differed substantially in how intentionally they coordinated grading-related instruction across coursework, modeled it, and connected it to field experiences. These differences shaped how pre-service teachers encountered, interpreted, and applied grading practices during their preparation, reflecting variation in program-level expectations, faculty coordination, and the extent to which grading was positioned as a shared instructional responsibility.

Within-case analysis first identified variation in how grading was discussed, emphasized, and enacted inside individual programs. Subsequent cross-case analysis revealed recurring patterns related to programmatic coherence, faculty coordination, and connections between coursework and clinical experiences. Analysis identified three distinct approaches to grading preparation across cases.

These approaches are presented as three themes representing distinct configurations of how grading and assessment instruction was structured within programs: fragmented passive exposure, emerging integration, and aligned programmatic approach. Together, these themes informed the development of the Continuum of Grading Approach Integration, which is presented in the cross-case synthesis. The following sections present each theme in turn, supported by faculty perspectives and programmatic artifacts, and then return to the continuum as a cross-case synthesis illustrating how these approaches relate to one another.

Theme 1: Fragmented Passive Exposure

Programs situated at the *fragmented passive exposure level* of the continuum addressed grading in ways that were incidental, unstructured, and largely dependent on individual faculty members' initiative rather than on programmatic design. This pattern reflects the absence of shared program expectations, coordinated curriculum planning, and explicit responsibility for grading instruction across faculty. In this context, programmatic integration was minimal, with grading instruction occurring inconsistently across courses and without coordination among faculty or alignment with field experiences. In these contexts, grading was not positioned as a central component of teacher preparation nor explicitly connected to the philosophical tenets of SBG. Instead, exposure to grading practices occurred sporadically, often emerging only when prompted by student questions or field-based experiences. As a result, candidates encountered grading as a

peripheral topic rather than as a core instructional and ethical responsibility because no programmatic structures positioned grading as a shared focus across coursework.

Faculty participants frequently described relying on traditional grading practices (e.g., letter- and percentage-based grades) shaped by their own experiences in PK–12 or higher education, rather than by contemporary assessment research. In these contexts, grading was not positioned as a program-level priority, which limited opportunities for faculty collaboration or intentional integration across coursework. One faculty member at Hackberry College characterized grading instruction as incidental, explaining that it was "a quick aside" during a lesson planning unit and noting, "It's something we talk about when students ask, but I wouldn't say it's a focus." This reactive approach to grading instruction suggests that, in the absence of programmatic structures, grading becomes contingent on individual faculty decisions rather than on coordinated curricular intent. Overall, this framing reflects a reactive approach to grading instruction, in which assessment literacy is treated as secondary to other pedagogical priorities.

Across cases, the lack of intentional curricular attention to grading was often justified by an assumption that assessment knowledge would be acquired during student teaching. Faculty deferred responsibility to cooperating teachers, despite recognizing inconsistencies in field-based practices. A reliance on field placements appeared to function as an implicit substitute for programmatic instruction, despite acknowledged inconsistencies across clinical contexts. As Lew, a faculty member from Sumac College, acknowledged, "We kind of assume they'll see grading during student teaching, but whether that aligns with anything we've talked about is another story." Such comments illustrated how grading preparation was displaced from coursework to clinical contexts without guarantees of coherence or alignment.

The absence of shared language, coordinated faculty development, or institutional expectations around grading also resulted in fragmented learning experiences for teacher candidates. Although faculty expressed concern about inequitable grading practices in PK–12 schools, these concerns were not translated into systematic preparation. At this level of the continuum, grading remained disconnected from broader conversations about equity, mastery, and instructional decision-making. This reflects an absence of structures that would otherwise position grading as a shared programmatic responsibility. Consequently, it limits candidates' opportunities to critically examine how assessment practices shape student outcomes.

Theme 2: Emerging Integration

Programs classified at the *emerging integration level* demonstrated growing awareness of the importance of grading and assessment preparation, with evidence of faculty experimentation and limited curricular integration of SBG principles. Unlike fragmented contexts, these programs addressed grading more explicitly; however, integration remained partial, uneven, and highly dependent on individual faculty members, reflecting limited program-level structures to support coordination across courses. In this context, programmatic integration was partial, with grading instruction appearing in multiple courses but lacking consistent coordination, shared expectations, or systematic reinforcement across the program.

In these settings, grading instruction often appeared in isolated courses, seminars, or assignments. Faculty described introducing elements of SBG, such as rubric-based assessment, standards-referenced feedback, or reassessment opportunities, but without sustained scaffolding across the program. One participant, Anna from Cottonwood University, described how conversations about assessment had begun to surface at the departmental level, stating, "We've started talking more about assessment at department meetings, but it's still not something we've structured as a faculty." This sentiment reflected growing recognition of assessment's importance without corresponding institutional mechanisms to support coherence, resulting in uneven implementation across coursework and field experiences.

Faculty participants also articulated tensions between theoretical instruction and practical realities, particularly in field placements. While coursework emphasized equity and mastery, candidates often observed traditional grading practices during student teaching. Jane from Hackberry College captured this disconnect, noting, "We talk about equity and mastery in class, but then they go into a school that still uses traditional grading, and it confuses them." These disconnects complicated candidates' efforts to reconcile coursework with practice, highlighting how inconsistent program structures and field experiences shaped their understanding of grading.

Despite these challenges, emerging programs displayed notable potential for future alignment. Faculty were reflective and increasingly attuned to the ethical and instructional implications of grading. Participants frequently expressed enthusiasm for SBG's capacity to provide meaningful feedback and reduce subjectivity. As Lew from Sumac College explained, "Standards-based grading is far more specific than just slapping a B on a paper and handing it to a student . . . there's context there." Such perspectives suggest that while integration was incomplete, foundational beliefs necessary for systemic change were beginning to coalesce.

Theme 3: Aligned Programmatic Approach

Programs situated at the *aligned programmatic approach level* of the continuum demonstrated intentional, cohesive integration of SBG across coursework, clinical experiences, and faculty practice. In these contexts, grading was treated as a pedagogical and philosophical priority rather than as a discrete instructional topic because program structures intentionally reinforced shared expectations across coursework and field experiences. In this context, programmatic integration was intentional and sustained, with grading instruction coordinated across coursework, reinforced by faculty, and aligned with field experiences. Faculty shared a common language, curricular structures reinforced consistent expectations, and candidates encountered SBG as an integrated framework for thinking about teaching, learning, and equity.

A defining feature of aligned programs was explicit curricular scaffolding. SBG concepts were introduced in foundational coursework, expanded in methods classes, and applied during practicum and student teaching. Faculty described modeling grading practices using shared rubrics, performance scales, and transparent feedback systems. Neice from Sumac College emphasized the importance of consistency, stating, "This scale is the same across all the courses in the teacher education program." Such alignment ensured that candidates repeatedly engaged with common assessment principles, which strengthened their ability to interpret and apply grading practices consistently across contexts.

Aligned programs also benefited from institutional structures that supported a unified approach. Faculty collaboration shared philosophical commitments, and program-wide conversations about grading were common. Arthur from Cottonwood University described a program-level effort to strengthen alignment, explaining that the goal was to evidence candidate growth toward mastery of professional standards and to provide opportunities for reflection and revision. Arthur emphasized, "How are you evidencing that your pre-service teachers are moving toward proficiency and mastery of the standards, and how are they having the opportunity to come back and talk about their growth?" This focus positioned grading as integral to professional learning rather than as a terminal judgment.

Connections between coursework and field experiences were particularly strong in aligned programs. Faculty worked intentionally with clinical partners to ensure candidates observed and practiced grading approaches consistent with program values. When discrepancies arose, they were addressed through reflective dialogue rather than ignored. This integration helped candidates navigate real-world constraints while maintaining a commitment to equitable assessment, as program structures provided consistent frameworks for interpreting grading practices. As Patricia from Hackberry College noted when discussing rubric-based grading, "Here are my expectations. This is a way to meet it. But if you want to go above and beyond, here's how." Such practices reinforced grading as a communicative and instructional tool, preparing candidates to enact SBG principles within diverse PK–12 contexts.

Cross-Case Synthesis: Continuum of Grading Approach Integration

Synthesizing findings across the three themes, the Continuum of Grading Approach Integration (see Figure 2) was developed to represent patterned variation in how undergraduate TEPs prepared pre-service teachers for SBG. The continuum reflects cross-case patterns in the degree of programmatic integration, ranging from minimal and fragmented to intentional and sustained coordination across teacher preparation programs. Each position on the continuum captures a qualitatively distinct approach to grading instruction, as observed across cases and grounded in participant perspectives and programmatic artifacts.

At one end of the continuum, fragmented passive exposure represents programs in which grading instruction was present but inconsistently integrated, leaving candidates to navigate grading practices largely through individual courses or practicum experiences. Emerging integration reflects increased intentionality and faculty awareness of grading reform, with grading instruction embedded across multiple courses but lacking formalized program-wide coordination. At the other end of the continuum, an aligned programmatic approach represents contexts in which grading and assessment instruction was systematically coordinated across coursework and field experiences, supported by shared language, modeling of practices, and structured reflection.

The continuum is not intended to suggest a linear progression or developmental sequence that programs must follow, nor does it function as an evaluative rubric. Instead, it serves as an analytic heuristic for understanding variation across teacher-preparation contexts and for supporting reflection among teacher-education faculty and program leaders. Programs may exhibit characteristics of multiple positions on the continuum simultaneously, and placement may shift over time in response to institutional, policy, or personnel changes. As such, the continuum offers a flexible framework for interpreting how grading-related instruction is structured within TEPs rather than a prescriptive model for reform.

Figure 2

A Continuum of Grading Approach Integration in Undergraduate Teacher Education Programs

Programmatic Integration of Grading Instruction		
Fragmented Passive Exposure	**Emerging Integration**	**Aligned Programmatic Approach**
• Grading instruction appears inconsistently across coursework • Limited coordination among faculty regarding grading practices • Exposure to standards-based grading depends largely on practicum placements • Traditional grading practices dominate teacher education coursework	• Grading instruction embedded across multiple courses • Growing attention to standards-based grading principles • Informal faculty collaboration around grading and assessment • Limited program-level coordination or shared expectations	• Grading and assessment instruction coordinated across coursework • Shared faculty language and expectations related to grading • Coursework models standards-aligned assessment and feedback • Field experiences intentionally connected to programmatic frameworks
Grading Practices Addressed Across Coursework and Field Experiences		

Note. The continuum represents analytically distinct approaches to programmatic integration of grading instruction observed across cases and does not imply a linear progression or evaluative hierarchy. Programs may exhibit characteristics of multiple positions simultaneously.

DISCUSSION

This study shifts attention from individual grading practices to the organizational conditions that shape how grading is learned within teacher preparation programs. Rather than functioning as a discrete instructional topic, grading emerged as a program-level structure shaped by the degree of coordination across coursework, faculty practices, and field experiences. These findings suggest that pre-service teachers' understanding of grading is less a function of isolated instructional exposure and more a product of how programs structure opportunities to engage with assessment as part of instructional decision-making. These findings suggest that grading practices are not simply adopted by individual teachers but are socially and structurally constructed through the organization of teacher preparation programs, extending existing conceptions of grading as an individual instructional practice to a programmatic and organizational phenomenon.

These findings extend prior research on grading reform, which has largely focused on implementation challenges in PK–12 contexts (Brookhart et al., 2016; Townsley & Buckmiller, 2020), by demonstrating that coherence in grading practices is shaped well before teachers enter the classroom. While existing literature emphasizes the role of teacher beliefs and professional learning, the findings in this study highlight the importance of programmatic structures in shaping how those beliefs are formed and enacted. In this way, grading reform can be understood not only as a school-based initiative but as a function of how preparation programs organize learning opportunities across the teacher development pipeline. This reframing positions teacher preparation as a critical, yet under-theorized, site of grading reform, suggesting that coherence in grading practices is not only enacted in schools but also constructed through the design of preparation programs.

The Continuum of Grading Approach Integration offers a conceptual contribution by theorizing variation in grading preparation as differences in the findings rather than differences in individual exposure. By conceptualizing grading preparation along a continuum, the study advances a framework for understanding how organizational structures within teacher education shape candidates' opportunities to learn, interpret, and enact grading practices.

Collectively, this study contributes to grading reform scholarship by extending theoretical attention from classroom practice to the organizational structures of teacher preparation. While existing literature has emphasized the role of teacher beliefs, policy contexts, and professional learning, the findings suggest that programmatic integration is a key mechanism through which grading practices are learned and normalized. By foregrounding coherence across coursework, faculty practice, and field experiences, this study offers a theoretical lens for understanding how preparation environments shape the future enactment of grading practices in PK–12 education.

IMPLICATIONS

Together, the findings suggest that preparation for SBG is shaped less by isolated instructional choices and more by how grading and assessment instruction is positioned across teacher education programs as a whole. Viewing grading preparation through a programmatic lens foregrounds the importance of alignment and intentional design in shaping pre-service teachers' readiness to engage with grading reform. The following implications consider how teacher education programs may respond to these findings to support sustained, context-sensitive preparation for evolving PK–12 grading practices.

Implications for Teacher Education

The findings of this study suggest how grading and assessment instruction structured within TEPs may shape pre-service teachers' readiness to engage in SBG practices encountered in PK–12 schools. Rather than emphasizing individual faculty adoption of specific grading strategies, the Continuum of Grading Approach Integration (see Figure 2) highlights the importance of a coherent programmatic structure in how grading principles are introduced, reinforced, and modeled across coursework and field experiences. For TEPs, this suggests that grading preparation may be strengthened by attending to coordination across courses and instructional contexts rather than relying on isolated or course-specific exposure.

One implication for teacher education faculty is the importance of shared language and expectations regarding grading and assessment. Participants in programs characterized as having stronger alignment described faculty collaboration on how grading concepts were framed, which supported candidates in developing a more coherent understanding of grading as part of instructional decision-making. While institutional constraints may limit the extent to which SBG can be fully implemented within university courses, modeling key principles such as transparent success criteria, standards-aligned assessment design, and formative feedback may help bridge the gap between theory and practice for pre-service teachers.

The findings also point to the role of structured reflection in connecting coursework and field experiences. In aligned programmatic contexts, faculty intentionally supported candidates in analyzing grading practices observed in PK–12 placements using shared frameworks and program-level language. This approach appeared to support candidates in navigating variation across school contexts while maintaining conceptual alignment. TEPs may benefit from designing reflective opportunities that explicitly connect grading practices encountered in practicum settings to concepts introduced in coursework.

Finally, the continuum may serve as a reflective tool for TEPs seeking to examine how grading-related instruction is currently positioned within their curricula. Rather than functioning as a prescriptive model, the continuum offers faculty and program leaders a way to consider questions of coherence and intentionality in grading preparation. Used in this way, the continuum may support program-level dialogue about how teacher preparation can respond to ongoing grading reforms in PK–12 education while remaining responsive to institutional context and constraints.

Implications for Educational Leadership

The findings of this study emphasize the role of educational leadership in shaping the conditions under which grading reform is understood, supported, and sustained across the teacher preparation pipeline. As SBG continues to expand in PK–12 contexts, school and system leaders increasingly rely on newly licensed teachers to enter classrooms with a foundational understanding of grading principles and practices. The Continuum of Grading Approach Integration (see Figure 2) suggests that variation in teacher preparation may influence how novice teachers engage with grading reform initiatives upon entering the profession.

For educational leaders working in PK–12 systems, these findings highlight the importance of communication and partnership with teacher preparation programs. District leaders implementing or refining SBG practices may benefit from articulating expectations for grading and assessment to inform clinical partnerships, student teaching placements, and mentoring structures. While TEPs operate independently of PK–12 districts, intentional collaboration may help ensure greater consistency between how grading is discussed in preparation programs and how it is implemented in schools.

At the postsecondary level, academic leaders and program administrators play a critical role in fostering conditions for programmatic alignment. The findings suggest that grading preparation is influenced not only by individual faculty expertise but also by institutional structures such as curriculum design processes, program review cycles, and faculty collaboration norms. Educational leaders within higher education may use the continuum as a reflective tool to examine how grading-related instruction is positioned across programs and to support dialogue around modeling and alignment without prescribing uniform practices.

Limitations

As with all qualitative case study research, the findings of this study should be interpreted considering several limitations. First, the study examined three small private undergraduate TEPs within a single Midwestern state. While this shared regional and policy context supported analytic comparison, the findings are not intended to be statistically generalizable to all teacher preparation programs. Instead, the study offers analytic insights into how grading and assessment instruction may be structured within similar institutional contexts.

Second, the study focused on interview data and programmatic documents rather than direct observation of coursework or analysis of pre-service teachers' instructional practices. Although faculty interviews and document analysis provided insight into how grading instruction was positioned and coordinated within programs, future research could extend these findings by examining how pre-service teachers experience and enact grading practices during coursework and clinical placements.

Third, the study centered on teacher preparation rather than on PK–12 implementation of SBG. While the findings are situated within broader grading reform efforts, the study does not evaluate the effectiveness of SBG practices in schools or the extent to which preparation experiences translate into classroom practice. As such, claims regarding PK–12 outcomes are necessarily limited.

Finally, institutional constraints within higher education, including grading policies and program requirements, may have shaped how faculty described and enacted grading practices. These contextual factors are an important consideration when interpreting findings and underscore the need to view grading preparation within the broader organizational environments in which TEPs operate.

Future Research

The findings of this study suggest several directions for future research on grading reform and teacher preparation. First, additional studies could examine how pre-service teachers interpret and apply grading-related instruction as they transition from coursework to clinical placements and early-career teaching. Longitudinal research following candidates from preparation programs into PK–12 classrooms would help illuminate how programmatic coherence in grading preparation influences novice teachers' grading practices over time.

Future research could also extend this work by examining a broader range of institutional contexts. Studies that include public universities, alternative licensure pathways, or institutions operating in different policy environments may provide additional insight into how grading preparation is shaped by organizational structures, accreditation requirements, and state-level mandates. Comparative studies across regions or states could further clarify how local grading reform efforts intersect with teacher preparation practices.

In addition, research incorporating multiple stakeholder perspectives would strengthen understanding of grading preparation as a system-level phenomenon. Including pre-service teachers, mentor teachers, and PK–12 administrators alongside teacher education faculty could offer a more comprehensive view of how grading expectations are communicated, reinforced, and negotiated across preparation and induction contexts. Mixed-methods approaches may be particularly useful in examining how perceptions of grading preparation align with observed practices.

Finally, future scholarship could explore how the Continuum of Grading Approach Integration (see Figure 2) functions as a reflective or analytic tool within TEPs. Research examining how faculty and program leaders use the continuum to guide curriculum review, professional dialogue, or program redesign could help clarify its utility and limitations. Such work would contribute to a growing body of research focused on integration and sustainability in grading reform across educational systems.

CONCLUSION

As SBG continues to shape assessment practices in PK–12 education, questions of how teachers are prepared to engage with these systems remain central to the sustainability of grading reform. This study contributes to that conversation by examining how undergraduate TEPs integrate grading and assessment instruction within their curricula. Through cross-case analysis, the study highlights patterned variation in how grading preparation is structured and coordinated, shifting attention from individual instructional practices to programmatic alignment.

The Continuum of Grading Approach Integration (see Figure 2) offers a conceptual contribution that captures this variation and provides a language for discussing how teacher preparation programs may differ in their approaches to grading instruction. Rather than positioning grading reform as a technical skill to be mastered in isolation, the findings suggest that

consistency across coursework, faculty practices, and field experiences plays a meaningful role in how pre-service teachers make sense of grading as part of instructional decision-making.

By centering teacher preparation within broader grading reform efforts, this study underscores the importance of alignment across educational systems. As PK–12 schools continue to reconsider how grading communicates learning, TEPs and educational leaders alike are positioned to reflect on how preparation structures support or constrain coherence across teacher preparation programming. Viewed in this way, moving from fragmented to aligned approaches to grading preparation becomes not a prescriptive pathway, but an ongoing, context-sensitive process that shapes the future enactment of grading practices in schools.

REFERENCES

Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. *Journal of Teacher Education, 36*(4), 46–49. https://doi.org/10.1177/002248718503600412

Battistone, W., Buckmiller, T., & Peters, R. (2019). Assessing assessment literacy: Are new teachers prepared to assume jobs in school districts engaging in grading and assessment reform efforts? *Studies in Educational Evaluation, 62*(1), 10–17.https://doi.org/10.1016/j.stueduc.2019.04.009

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership, 69*(3), 1–5. https://www.ascd.org/el/articles/starting-the-conversation-about-grading

Brookhart, S. M. (2015). Graded achievement, tested achievement, and validity. *Educational Assessment, 20*(4), 268–296. https://doi.org/10.1080/10627197.2015.1093928

Brookhart, S. M. (2017). *How to create and use rubrics for formative assessment and grading.* ASCD.

Brookhart, S. M., Guskey, T. M., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803–848. https://doi.org/10.3102/0034654316672069

Crotty, M. (2015). *The foundations of social research: Meaning and perspective in the research process* (Reprinted.). Sage. (Original work published 1998)

Feldman, J. (2019). Beyond standards-based grading: Why equity must be part of grading reform. *Phi Delta Kappan, 100*(8), 52–55. https://kappanonline.org/standards-based-grading-equity-reform-feldman/

Fordham Institute. (2025). "Equitable" grading through the eyes of teachers. https://fordhaminstitute.org/national/research/equitable-grading-through-eyes-teachers

Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership, 59*(1), 20–27. https://uknowledge.uky.edu/edp_facpub/8

Guskey, T. R. (2013). The case against percentage grades. *Educational Leadership, 71*(1), 68–72. https://www.ascd.org/el/articles/the-case-against-percentage-grades

Guskey, T. R. (2015). *On your mark: Challenging the conventions of grading and reporting.* Solution Tree Press.

Guskey, T. R., & Jung, L. A. (2013). *Answers to essential questions about standards, assessments, grading, and reporting.* Corwin.

Guskey, T. R., Townsley, M., & Buckmiller, T. M. (2020). The impact of standards-based learning: Tracking high school students' transition to the university. *NASSP Bulletin, 104*(4), 257–269. https://doi.org/10.1177/0192636520975862

Kohn, A. (2011). The case against grades. *Educational Leadership, 69*(3), 28–33. https://www.ascd.org/el/articles/the-case-against-grades

Lewis, D. (2022). Impacts of standards-based grading on students' mindset and test anxiety. *Journal of the Scholarship of Teaching and Learning, 22*(2), 67–77. https://doi.org/10.14434/josotl.v22i2.31308

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Sage.

Link, L. J. (2019). Teachers' perceptions of grading practices: How pre-service training makes a difference. *Journal of Research in Education, 28*(1), 62–91.https://eric.ed.gov/?id=EJ1168160

Link, L. J., & Guskey, T. R. (2022). Is standards-based grading effective? *Theory Into Practice, 61*(4), 406–417. https://doi.org/10.1080/00405841.2022.2107338

Marsh, V. L. (2023). Standards-based grading: History, practices, benefits, challenges, and next steps. *University of Rochester Warner School of Education, 1*(1), 1–22. https://www.rochester.edu/warner/cues/wp-content/uploads/2023/11/Standards-Based-Grading.pdf

Merriam, S. B. (1988). *Case study research in education: A qualitative approach.* Jossey-Bass.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (4th ed.). Sage.

Morris, S. R., Maranto, R. A., & McKenzie, S. E. (2023). Examining teacher support for grading equity. *Journal of Research in Education, 32(*2), 1–42. https://files.eric.ed.gov/fulltext/EJ1412057.pdf

Morris, S. R., Parra-Martinez, A., Wai, J., & Maranto, R. (2025). Administrators' perceptions of standards-based grading practices (EdWorkingPaper No. 24-917). Brown University. https://edworkingpapers.com/sites/default/files/ai24-917.pdf

National Center for Educational Statistics. (2022). College Navigator. Institute of Education Sciences. https://nces.ed.gov/collegenavigator/

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Schimmer, T., Hillman, G., & Stalets, M. (2018). *Standards-based learning in action.* Solution Tree Press.

Seidman, I. (2019). *Interviewing as qualitative research* (5th ed.). Teachers College Press.

Stake, R. E. (2006). *Multiple case study analysis*. Guilford Press.

Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles of standards-based grading. *The Educational Forum, 75*(3), 210–227.https://doi.org/10.1080/00131725.2011.577669

Townsley, M. (2020). Grading principles in pandemic-era learning: Recommendations and implications for secondary school leaders. *Journal of School Administration Research and Development, 5*(S1), 8–14. https://doi.org/10.32674/jsard.v5iS1.2760

Townsley, M., & Buckmiller, T. M. (2020). Losing As and Fs: What works for schools implementing standards-based grading? *Educational Considerations, 46*(1), 1–10. https://doi.org/10.4148/0146-9282.2204

Townsley, M., Buckmiller, T., & Cooper, S. (2019). Anticipating a second wave of standards-based grading implementation and understanding the barriers: Perceptions of high school principals. *NASSP Bulletin, 103*(4), 281–299. https://doi.org/10.1177/0192636519882084

Townsley, M., & Wilcox, J. (2024). Standards-based grading: Core principles and current implementation debates. *The Clearing House, 97*(3), 76–86. https://doi.org/10.1080/00098655.2024.2357615

Yin, R. K. (2018). Case study research and applications (6th ed.). Sage.

Zhang, B. (2024). A didactic explanation of standards-based grading. *Educational Research and Development Journal, 27*(2), 42–56. https://files.eric.ed.gov/fulltext/EJ1462852.pdf

KEEGAN VANDEVENDER, PhD, is a recent graduate of Drake University. He is a lead teacher and student success coach in the Oskaloosa Community School District, and an adjunct professor at Central College, Northwestern College, and William Penn University. His major research interests lie in the areas of standards-based grading, educational assessment, and change management models in higher education. Email: kjv.phd@gmail.com

Manuscript submitted: ***September 7, 2025***
Manuscript revised: ***December 28, 2025***
Manuscript revised: ***April 4, 2026***
Accepted for publication: ***April 14, 2026***

***Special Issue**: Exploring Current & Future Grading Practices in K-12 Schools*
Research Article

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 105-120
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

A Matter of Standards: Addressing Inconsistencies in Grading Practices in High School English, Mathematics, Science, and Social Studies Courses

Allie Kollman
TrueMarks Education, USA
Erin Sinner
Hannah Nurmi
Cailen O'Shea
North Dakota State University, USA

ABSTRACT

This study examines the alignment between high school students' summative assessment scores and their final course grade in the point-based grading system to identify patterns of grade inflation or deflation. Using a non-experimental quantitative design, we analyzed gradebook data from 2,494 students in four subject areas at three high schools in the same Midwestern school district. Paired *t* tests revealed statistically significant discrepancies, with one-third of students receiving grades at least five percentage points different from their cumulative summative score averages, and 36% experiencing at least one letter grade shift. Grade inflation was notably more common than deflation. These findings raise concerns about the fairness and accuracy of point-based grading systems and support the need for more consistent, standards-based approaches that prioritize defined learning outcomes, transparent assessment criteria, and the separation of academic achievement from behavioral compliance to ensure grades accurately reflect student proficiency.

***Keywords*:** evidence-based grading, grade inflation and deflation, grading reform

Historically, grades have shaped K-12 education's approach to measuring student achievement, influencing everything from parental expectations to college admissions and scholarship decisions (Brookhart, 2004; Guskey, 2009). Grades carry significant weight in shaping a student's academic trajectory, including honor roll, course pathways, access to advanced courses, and eligibility for extracurricular programs (Griffin & Townsley, 2021).

Research shows that high school grades are a better predictor of four-year college graduation than standardized test scores (Allensworth & Clark, 2020; Galla et al., 2019). Given the high-stakes role of grades, it is critical that they accurately reflect students' learning. A course grade should communicate the extent to which students have achieved the course objectives (Allen, 2005), providing a fair and valid summary of what a student knows and can do (Brookhart et al., 2016; Linn & Miller, 2005; Marzano & Heflebower, 2011).

In practice, however, ensuring accuracy and consistency in grading remains a challenge. Teacher-assigned grades are often a composite of academic performance and non-cognitive behaviors (Brookhart et al., 2016; Guskey & Link, 2018), such as homework completion, class participation, effort, and behavior (Griffin & Townsley, 2021). Teachers commonly incorporate not only summative assessment scores but also homework completion, class participation, effort, and behavior when determining final letter grades (Guskey & Link, 2018; McMillan, 2018). As a result, grades may reflect a mix of achievement and compliance, creating a "hodgepodge" of disparate elements that do not consistently indicate mastery of standards (Brookhart, 1991).

This study investigates the alignment between students' summative assessment scores and their final course grades to determine the extent of grade inflation or deflation in current grading practices. While previous research on grade inflation

has often compared grades or GPAs to standardized test results or college outcomes, there is limited analysis of grade accuracy at the classroom level (Brookhart et al., 2016; Griffin & Townsley, 2021; Pattison et al., 2013). This study addresses that gap by examining how final grades, expressed both as numerical percentages and as letter grades, differ from students' summative assessment averages. Because summative assessments are designed to evaluate how well students have mastered specific learning standards (Brookhart et al., 2016; Marzano & Heflebower, 2011), they serve as a meaningful baseline for judging the validity of final grades.

By comparing final reported grades to the evidence of summative proficiency, we can identify whether grades are being systematically inflated or deflated relative to actual achievement. Such findings have direct implications for grading reform efforts, as they can reveal patterns of inconsistency that compromise fairness and transparency in student evaluation.

The purpose of this study is to determine the degree to which final course letter grades diverge from students' summative assessment scores, thereby uncovering potential patterns of grade inflation or deflation. Our research question is as follows: *To what extent do final grades, measured both as percentages and as letter grades, differ from summative assessment scores?*

By investigating this alignment, the study seeks to assess the accuracy of current grading practices and provide actionable insights for school leaders and educators committed to advancing standards-based and equitable grading systems and norms.

REVIEW OF THE LITERATURE

In this literature review, we examine grading practices, inconsistencies, grade inflation and deflation, and concerns of equity. We begin with a review of overall grading practices and transition into the full review.

Starch and Elliott (1912) revealed wide inconsistencies in grading, with teachers assigning vastly different grades to identical student work. More than a century later, little has changed. Grading practices still vary significantly from one teacher to another, even within the same school or district (Brookhart, 2004; Guskey & Link, 2018; McMillan, 2001). This variability stems from long-standing traditions, limited formal teacher training in effective grading methods (Guskey, 2009), and inconsistent or absent district policies outlining grading expectations. As a result, individual teachers often create their own systems, some relying on total point accumulation models, others using weighted categories such as tests, homework, or participation, leading to inconsistent criteria for calculating grades (McMillan, 2001).

Persistent Inconsistencies in Grading Practices

A key inconsistency is the inclusion of behavioral, or non-cognitive tasks, such as homework-related items or in-progress monitored learning, in academic grades. Griffin and Townsley (2021) define "employability scores" as points awarded for classroom behaviors rooted in present-day expectations, such as participation, citizenship, and timeliness, which are identified in this study as non-cognitive behaviors. While these behaviors are often valued in point accumulation grading models, they risk distorting the academic meaning of a student's grade because they can inflate or deflate grades regardless of actual proficiency of standards (Griffin, 2020; Guskey, 2009; McMillan, 2001).

To preserve the integrity of achievement data, researchers recommend reporting non-cognitive factors separately from academic grades (Brookhart, 2004; Feldman, 2018; Griffin & Townsley, 2021; Guskey, 2009). Although behaviors such as work habits or citizenship are important indicators of student development, they should be documented independently to avoid confounding academic achievement (Link & Guskey, 2022).

Mixing non-academic elements with academic performance can yield grades that send misleading signals about what a student actually knows. In one study, homework and employability points in high school math grades resulted in significantly different outcomes compared to assessment scores alone (Griffin & Townsley, 2021). For example, a student with strong content knowledge might receive a low final grade due to missing assignments, while another student with modest understanding could receive a higher grade due to effort or extra credit (Guskey, 2009; Reeves, 2008). Such examples undermine the core purpose of grades as accurate communication of academic proficiency.

Distortions in Grade Meaning: Inflation and Deflation

These inconsistencies have raised growing concern about grade inflation and deflation. Grade inflation is defined as a rise in students' grades without a corresponding increase in achievement (Yeritsyan et al., 2022). Numerous studies have documented rising high school GPAs despite stagnant or declining standardized test scores (Camara et al., 2003; Godfrey, 2011; Woodruff & Ziomek, 2004). For example, one state analysis found that students with relatively low SAT scores were

nonetheless earning higher high school GPAs than in the past, suggesting inflated grades (Godfrey, 2011). Grade deflation can occur in the opposite scenario, when grades underrepresent students' true academic achievement (Brookhart et al., 2016; Guskey, 2009; Pattison et al., 2013). Some researchers note that focusing only on rising average GPAs can mask considerable variability, including cases where certain students receive lower grades than their level of proficiency (Pattison et al., 2013).

Both inflation and deflation highlight a disconnect between report card grades and demonstrated proficiency. For instance, a study in North Carolina found that Algebra I end-of-course exam scores were more predictive of ACT math performance than teacher-assigned course grades, implying that the standardized exam provided a more accurate gauge of content mastery than the teacher-assigned grade (Gershenson, 2018). Similarly, in a large urban high school, 43% of students' semester math grades differed by 5 or more percentage points from their summative averages, and 13% of students experienced inflation or deflation equivalent to an entire letter grade (Griffin & Townsley, 2021). This suggests a substantial number of students receive grades that do not accurately reflect their academic proficiency.

Point Accumulation Grading System and Standard-Based Grading Alternatives

These inconsistencies and misalignments highlight a clear problem: if final grades do not correspond to demonstrated learning, educational stakeholders may be misled about student performance and achievement. In response to these concerns, scholars and practitioners have increasingly turned to standards-based grading (SBG) and Evidence-Based Grading (EBG), both of which aim to improve grading accuracy and consistency by fundamentally restructuring how grades are constructed (Knight & Cooper, 2019; Muñoz & Guskey, 2015). In point accumulation systems, final grades are typically derived by aggregating points or scores across heterogeneous tasks (e.g., tests, homework, extra credit, participation), which combines academic achievement with non-academic factors. In contrast, SBG and EBG organize grades around clearly defined learning standards and rely on multiple sources of academic evidence to determine proficiency levels, with the goal of making grades more accurate reflections of what students know and can do (Reibel et al., 2025; Townsley, 2018). SBG, EBG, and similar reform models advocate for separating academic achievement from non-academic factors in grade reporting. For instance, teachers report work habits and behavior separately from academic performance so that the course grade itself represents only mastery or proficiency of the subject's standards (Brookhart et al., 2020; Link & Guskey, 2022; O'Connor, 2009). By doing so, schools hope to minimize the grade discrepancies caused by homework, extra credit, or behavior points and thereby improve the integrity and equity of grades; the final grade is intended to communicate what students know and can do in relation to specific standards, rather than the extent to which they complied with classroom expectations (Link & Guskey, 2022; O'Connor, 2009). Because summative assessments are designed to evaluate mastery of standards, they function as the primary evidence of proficiency within SBG and EBG frameworks. Comparing final course grades to summative assessment performance provides a lens for evaluating the extent to which points-based grading systems align, or fail to align, with evidence-based conceptions of grading accuracy and equity.

Equity and Bias in Grading Outcomes

Concerns about equity have come to the forefront of grading discussions. Traditional grading practices that include subjective components can inadvertently introduce bias, potentially disadvantageous to students from historically marginalized cultural or socioeconomic backgrounds (Feldman, 2018). When teachers factor in behavior or compliance that may be influenced by cultural norms or teacher expectations, grades can reflect implicit biases rather than purely academic performance. Assessment experts have concluded that integrating non-academic factors into grades mathematically introduces additional subjectivity and potential bias (Guskey, 1994; Ornstein, 1994). Ensuring that grades are an accurate and bias-free representation of learning is thus not only a matter of better measurement, but also a matter of justice, so that students from all backgrounds have equal opportunities to succeed based on their actual learning. The following section addresses the methodology used to address the research question guiding this study.

METHODOLOGY

This study employed a non-experimental quantitative research design to examine grade inflation and deflation across three public high schools. This approach was selected because it enabled the objective analysis of existing numerical data without manipulating variables or introducing interventions. A non-experimental design is appropriate for studies that seek to explore relationships or patterns using existing data without manipulation (Creswell & Creswell, 2018). The study focused

on identifying discrepancies between students' average summative assessment scores and final course grades, both percentages and letter grades, to detect patterns of grade inflation or deflation within real-world classroom grading systems.

The use of digital gradebook data allowed for a direct and systematic comparison between students' final course grades and underlying measures of summative performance, non-cognitive behaviors, and formative assignments such as homework. Participants were drawn from three high schools within a single public school district in the Midwestern United States. This district was purposefully selected due to its diverse student population, its ongoing transition to evidence-based grading practices at the secondary level, and its expressed commitment to improving grading consistency. Additionally, the district's centralized access to electronic gradebook systems and its active engagement in broader grading reform efforts made it a strategically relevant site for studying patterns of grade inflation and deflation. District and building-level approval was obtained in accordance with the district's research policy. To protect privacy, strict confidentiality protocols were followed throughout the study: no individual student or teacher could be identified, and all findings are reported in aggregate. The research protocol was reviewed and approved by the appropriate authorities, and all procedures adhered to ethical guidelines for the use of student data, including compliance with FERPA standards for de-identification and data security.

The final sample included 48 high school teachers (16 per school) and their students. At each school, four teachers were randomly selected from four core subject areas: English, mathematics, science, and social studies. This approach ensured disciplinary representation and reduced selection bias. These teachers' classes yielded a combined student sample of 2,558, though 64 students were excluded due to incomplete summative data, resulting in a final student sample of $N = 2{,}494$. Students with incomplete summative data were excluded from the analysis. Students with incomplete summative data ($n = 64$) were excluded because partial summative records (e.g., only one early summative score with later assessments blank) could produce misleading summative averages and biased paired comparisons with final course grades. Many of these 64 students transferred into the district late in the marking period. This sample size is consistent with comparable peer-reviewed studies on grading practices and is considered sufficient for generating statistically significant findings (Griffin, 2020; Griffin & Townsley, 2021; Pollio & Hochbein, 2015).

Data Collection

Gradebook data were collected for each selected teacher following institutional and district-level approvals. After receiving permission from the school district, the research team obtained digital copies of the teachers' gradebooks from the district's electronic grading system. To protect privacy, all student and teacher identifiers were anonymized prior to data access. This included the removal of names, student ID numbers, and teacher identifiers, which were replaced by randomized numerical codes.

Each gradebook contained detailed records of student performance, including individual assignment scores, grading category weights, and final course grades. From these records, key data points were extracted and organized for analysis: each student's final letter grade, final overall percentage, average summative assessment scores, and non-cognitive behaviors (e.g., homework and non-academic factors). A summative average was calculated for each student by averaging the scores of all summative assignment scores in that course.

Similar to the work of McMillan and Hellsten (2010), summative assessments in this study are defined as major evaluations of learning because they are intended to assess learning against course standards and learning objectives. In each teacher's gradebook, these assessments were identified as unit tests, projects, or other assessments designated as "summative." To identify these summative assessments consistently across gradebooks, a standardized extraction protocol was used based on the teacher's gradebook category labels (e.g., "Summative," "Tests," "Projects," "Exams") and assignment labels within the district's electronic grading system. Extracted data were reviewed and cross-checked for coding consistency and calculation accuracy prior to analysis. Accordingly, the analysis reflects summative designations as recorded in teachers' operational gradebook practices. See Table 1 for the example of summative assessment averages.

Once the data were cleaned and organized, the next step involved accounting for variations in grading systems across participating teachers. Some teachers used a total points system (unweighted), while others used weighted categories, such as a 70/30 or 60/40 split between summative assessments and non-cognitive behaviors. To ensure comparability across systems, summative scores were isolated from non-academic components in every case. For teachers using weighted categories, summative assessment percentages were taken directly from the gradebook's summative category calculation. For teachers using total points, summative assignments were identified by their labels or gradebook categorization (e.g., "Exam" or "Test"), and the student's summative percentage was computed as points earned out of points possible on those summative assessments.

Each student's final course percentage and final letter grade were recorded alongside their summative average percentage and its corresponding letter grade equivalent. All data were cross-checked for accuracy (e.g., ensuring letter grades correspond to the reported percentage range). The complete dataset was then entered into SPSS (Version 28) for statistical analysis.

Table 1

Example of summative assessment averages from the point accumulation model

Assessment	Student 1	Student 2
Homework (%)	100	65
Participation (%)	95	60
Homework (%)	100	70
Formative Quiz (%)	92	75
Summative #1 (%)	78	92
Homework (%)	100	80
Summative #2 (%)	72	86
Participation (%)	90	50
Summative #3 (%)	64	94
Summative Avg (%)	71.33	90.66
Final Course Grade	B	C

Note. Student 1's higher final course grade (B), despite a lower summative average, reflects the influence of non-cognitive factors (e.g., formative assessments, homework, participation) in a point-accumulation grading model.

Statistical Analysis

To address the research questions, we conducted a series of quantitative analyses comparing students' summative performance to their final grades. For each student, the summative assessment average (expressed as a percentage) was directly compared to the final course percentage. This enabled the calculation of the difference, which effectively measured grade inflation or deflation. A positive difference (i.e., a final grade higher than the summative average) was interpreted as grade inflation, whereas a negative difference was interpreted as grade deflation.

Next, to examine potential discrepancies at the categorical level, we converted both the summative assessment average and the final letter grade into a standardized 4.0 scale (as a categorical value, A = 4, B = 3, C = 2, D = 1, F = 0). This transformation allowed for direct comparison of letter grades in numeric terms (A= 4, B = 3, C = 2, D = 1, F = 0), enabling us to detect shifts equivalent to full or partial letter grade changes. The district used a common letter-grade cut-score scheme across courses (e.g., 90–100 = A, 80–89 = B, 70–79 = C, 60–69 = D, below 60 = F), so formal syllabus review was not required to establish the cut-score framework used for the analytic letter-grade conversion.

We conducted two primary statistical tests using IBM SPSS Statistics (Version 28), as these tests can detect significant differences between two related sets of data: students' summative assessment average and their final course grades. The first test compared the numeric differences between final letter grade percentage and the summative assessment percentage to determine whether grades were inflated or deflated. The second test compared the summative and final letter grades after converting them to a standard numerical scale (A = 4, B = 3, C = 2, D = 1, F = 0) to examine whether similar patterns emerged when grades were treated categorically. Both these analyses allowed for patterns to be identified in grading consistency and accuracy between assessed performance and reported grades.

Paired-Samples t-Test on Percentage Grades

The study performed a paired-samples *t*-test comparing the mean of students' final percentage grades to the mean of all the students' summative assessment percentage scores (Griffin & Townsley, 2021; Pollio & Hochbein, 2015). Each student's summative average was calculated by averaging all summative assessments recorded in the course gradebook. Each student contributed a pair of values: final letter grade percentage and summative average percentage. This test determines whether, on average, final letter grades differed significantly from what purely summative scores would predict.

The paired *t*-test was used because the research question focuses on detecting inflation and deflation between two related measures for the same individuals. The analysis focused on within-student differences between two paired measures (summative average and final course grade); therefore, multilevel modeling was not used. This approach has been used in prior grading research to evaluate alignment between assessment-based performance and final grades (Griffin & Townsley, 2021). The analysis was two-tailed and used $\alpha = .05$, a conventional threshold for significance that allows for the detection of both grade inflation and deflation, rather than assuming a directional hypothesis.

Paired-Samples *t*-Test on Letter Grades

To address the research questions, we conducted a paired *t*-test on categorical letter grades to examine discrepancies in final letter grades across grade categories. Each student's final letter grade and the letter grade corresponding to their summative average were converted into numeric values on a standard 4.0 scale (A = 4, B = 3, C = 2, D = 1, F = 0). For example, if a student's summative average was 78%, it was treated as a "C" (2.0) for comparison to the final letter grade ("B" = 3.0). This approach follows Griffin and Townsley (2021) and Pollio and Hochbein (2015), who similarly used numeric conversions of letter grades for analysis. Using these numeric equivalents, a paired *t*-test compared the mean final letter-grade points with the mean summative-assessment points for the sample. This test provides insight into whether students tended to receive higher or lower letter grades than their level of proficiency would recommend.

While letter grades are ordinal, this numerical translation allowed us to apply parametric analysis to gauge average differences in letter grade (with 1.0 indicating a one-letter grade difference or with 2.0 indicating a two-letter grade difference). Both tests were run on the full sample of 2,494 students, yielding corresponding *p*-values and effect size estimates. This is significant because it quantifies whether students are receiving grades that accurately reflect their demonstrated academic proficiency based on summative assessments. Understanding these differences is essential for evaluating the integrity of grading practices and identifying systemic inflation or deflation patterns across schools or teachers.

To quantify the extent of grade inflation and deflation in practical terms, we also calculated the frequency of discrepancies beyond certain thresholds. Following the approach of Griffin and Townsley (2021), we identified each student who had a notable gap between their summative performance and final grade. Two percentage-based thresholds were used: 5% and 10%. If a student's final percentage exceeded their summative average by 5% or more, we classified that as a case of grade inflation; if it fell 5% or more, it was classified as grade deflation. A difference of five percent corresponds roughly to half a letter grade in typical grading scales (Guskey, 2013; Schinske & Tanner, 2017; Schneider & Hutt, 2014). More stringently, a 10% or greater difference was recorded as a case of major inflation or deflation, approximately equivalent to a full letter grade discrepancy. For example, a student with an 80% summative assessment average who received a 90% final grade would be counted in the "≥10% inflation" group, while one with 80% summative but a 70% final would be "≥10% deflation."

We tallied the number and percentage of students falling into each category of inflation or deflation. Similarly, we examined categorical grade shifts. We noted how many students' final letter grades were at least one or two full letters higher or lower than the letter grade indicated by their summative score. All metrics (5%/10% thresholds and 1-letter/2-letter discrepancies) were computed to supplement the *t*-test results, providing a clear picture of the prevalence of grade inflation/deflation in the sample.

Building on the foundational work of Cohen (1988) on effect size interpretation, effect sizes were calculated to assess the magnitude of any observed differences. Cohen's *d* was computed for each paired comparison (percentage and letter grade) by dividing the mean difference by the standard deviation of the difference scores (Borenstein et al., 2009; Cohen, 1988; Lakens, 2013). This provides a standardized measure of how large the grade inflation/deflation effect was. We interpreted *d* using conventional benchmarks: ~0.20 was considered a small effect, ~0.50 a medium effect, and ~0.80 or above a large effect (Cohen, 1988; Cohen, 1992; Lakens, 2013; Sullivan & Feinn, 2012). All analyses were conducted at the 95% confidence level, and results are presented with appropriate statistical notations (e.g., mean differences, *t* values, degrees of freedom, *p* values, and confidence intervals) (American Psychological Association, 2020; Wasserstein & Lazar, 2016; Wilkinson et al., 1999). This analytical strategy directly aligns with the research questions, determining whether final grades differ significantly from summative assessment averages and quantifying how many students are affected by such differences and to what extent.

RESULTS

To address the research question, *"To what extent do final grades, measured both as percentages and as letter grades, differ from summative assessment scores, indicating grade inflation or deflation?*" two complementary analyses were conducted, as shown in Table 2. First, a paired-samples *t*-test was used to compare students' final course percentages to their average summative assessment percentages. Second, a paired-samples *t*-test was performed on converted letter grades, using a 4.0 GPA scale (A = 4, B = 3, etc.), to examine discrepancies between categorical summative grades and final reported grades. These analyses allowed for both numerical and categorical evaluation of grading discrepancies, offering an in-depth understanding of grade inflation or deflation at both the system and student levels. This dual approach provided a more comprehensive understanding of the extent to which grading misalignment was present, or absent, within the district's current practices.

Final Grade Percentage vs. Summative Assessment Percentage Grades

The percentage-based analysis compared students' final overall grade percentages to their summative assessment grade percentages. The analysis revealed significant differences ($t_{2494} = 14.71$, $p < .001$) when comparing the means between students' final grade percentages ($M = 83.15$, $SD = 13.16$) and their summative assessment percentage grades ($M = 81.40$, $SD = 14.34$), suggesting an average inflation of 1.75 percentage points. The 95% confidence interval for the mean difference [1.52, 1.99] indicated consistency with grade inflation. This finding is important because it demonstrates that final course grades, which are the most visible on transcripts and used for GPA calculations, are, on average, modestly inflated compared to students' summative assessment performance. This suggests that non-cognitive behaviors may be elevating reported achievement beyond what assessments alone indicate, prompting deeper reflection on whether current grading practices genuinely represent student proficiency or merely reward compliance and conduct, thereby distorting our understanding of academic achievement.

To further examine the alignment between final grades and summative assessment performance, a Pearson correlation coefficient was calculated. This tool was chosen because it measures the strength and direction of the linear relationship between two continuous variables, providing a clear indication of how closely students' overall course outcomes align with their performance on key assessments. The analysis revealed a very strong positive correlation between final grade percentages and summative assessment percentages ($r = .91$, $p < .001$) (Goodwin & Leech, 2006). This indicates that students who performed well on summative assessments generally received correspondingly high final grades, although those final grades were, on average, slightly inflated. The effect size for the mean difference was also calculated using Cohen's *d*. Cohen's *d* was selected because it provides a standardized measure of the magnitude of the difference between two groups, allowing for easier interpretation and comparison across studies. The resulting value was 0.295, indicating a small to moderate effect.

Final Letter Grades vs. Summative Assessment Letter Grades

In addition to understanding the inflation and deflation percentages, the researchers also analyzed letter grades. A similar pattern of inflation emerged, with differences that are more pronounced in practical terms, depending on how letter grades are configured and the thresholds that are predetermined to generate an A, B, C, D, or F. Furthermore, transcripts report final letter grades, which may or may not include final percentage grades. Each student's final letter grade (A = 4, B = 3, C = 2, D = 1, F = 0) was compared to the letter grade equivalent of their summative assessment average. There were statistically significant ($t_{2494} = 17.35$, $p < .001$) differences when comparing the means between students' final letter grades ($M = 2.96$, approximately a B) and their summative assessment letter grades ($M = 2.74$, approximately a C), suggesting an average inflation of 0.22 letter grades. The average difference of +0.22 on a letter grade scale indicates that, on average, final grades were about a quarter of a letter grade higher than what summative assessments alone would justify. This is significant because it highlights how even a fraction of a letter grade in categorical grade inflation can have meaningful implications for student transcripts, college admissions, and scholarship eligibility. A discrepancy of 0.22 on a 4.0 GPA scale, while seemingly minor, may alter a student's reported academic standing and influence external perceptions of their achievement.

To measure the effect size, Cohen's *d* was utilized again. The effect size for this difference was $d = 0.348$, indicating a moderate effect that was larger than the effect observed in the percentage-based analysis ($d = 0.295$) (Cohen, 1988). There was also a strong correlation between final letter grades and summative letter grades ($r = .855$, $p < .001$), suggesting that students who performed well on summative assessments generally received correspondingly higher letter grades (Goodwin & Leech, 2006). This correlation was somewhat weaker than the correlation observed with percentage grades, which was

expected since converting percentages into letter grades introduces categorical thresholds (A = 4, B = 3, C = 2, D = 1, F = 0). This means that even small percentage-based discrepancies can lead to meaningful shifts in categorical outcomes, reinforcing the impact of grade inflation. See Table 2 for the summary of descriptive and inferential statistics.

Prevalence of Grade Inflation and Deflation

While the average discrepancy between final grade and summative assessments was modest, a significant proportion of students experienced substantial individual-level differences, highlighting instances of grading inconsistency. For example, approximately one-third (30.83%), or 769 students, demonstrated statistical discrepancies, with their final grades differing by at least 5 percentage points from the summative assessment averages. Specifically, grade inflation was more prevalent, affecting 21.97% of students, or 548 students, compared to grade deflation, which affected 8.86% of students, or 221 students. The reasoning for a higher discrepancy with grade inflation could be attributed to non-cognitive factors often used in point accumulation models that become perceived reasons for student success (Wormeli, 2017). These include points awarded for peripheral components like compliance, participation, effort, extra credit, etc. These factors can artificially elevate overall grades independent of true academic proficiency. This suggests that a portion of students had final grades that were significantly higher or lower than the summative assessment averages, beyond typical measurement variability.

Table 2

Summary of Descriptive and Inferential Statistics for Paired-Samples Comparisons of Final Grades and Summative Assessment Scores

Comparison	Final Grade Mean	Summative Assessment Mean	Final Grade *SD*	Summative Assessment *SD*	Mean Difference	*t*	*p*	Cohen's *d*	95% Confidence Level
Final Grade Percentage vs. summative percentage grades	83.15	81.40	13.17	14.34	1.76	14.73	<.001	0.295	[1.52, 1.99]
Final letter grades vs. summative letter grades	2.96	2.74	1.14	1.23	.22	17.36	<.001	0.348	[0.20, 0.25]

Further analysis revealed that 10.79%, or 269 students, had even larger discrepancies, with 7.66%, or 191 students, experiencing an inflation rate of 10% or more, and 3.13%, or 78 students, experiencing a deflation rate of the same magnitude. This results in approximately one in nine students receiving a final grade that differs from their summative assessments by at least 10 percentage points.

These discrepancies equate to half- or full-letter grade shifts relative to students' summative assessment percentages. For instance, a student achieving a summative assessment average of 77% could potentially receive a final grade percentage as high as 87% or as low as 67% when additional non-cognitive behaviors are considered. This highlights the significant practical implications for accurately reporting and interpreting student academic grades. See Table 3 for the frequency and percentages of grade inflation and deflation. Then see Figure 1 for the grade discrepancies between final grade percentages and summative assessment percentages, and Figure 2 for the distribution of discrepancy scores across the full sample.

Letter Grade Discrepancies

When letter grade discrepancies were analyzed categorically, the prevalence of grading inconsistencies was further highlighted. A total of 36.25%, or 904 students, received a final letter grade, differing by at least one full letter from their summative assessment letter grades. Grade inflation remained significantly more common, affecting 28.31% of students (706 students), compared to grade deflation, which impacted 7.94% of students (98 students). This indicates that nearly one-third of students ended the semester with higher letter grades than their summative performance suggested.

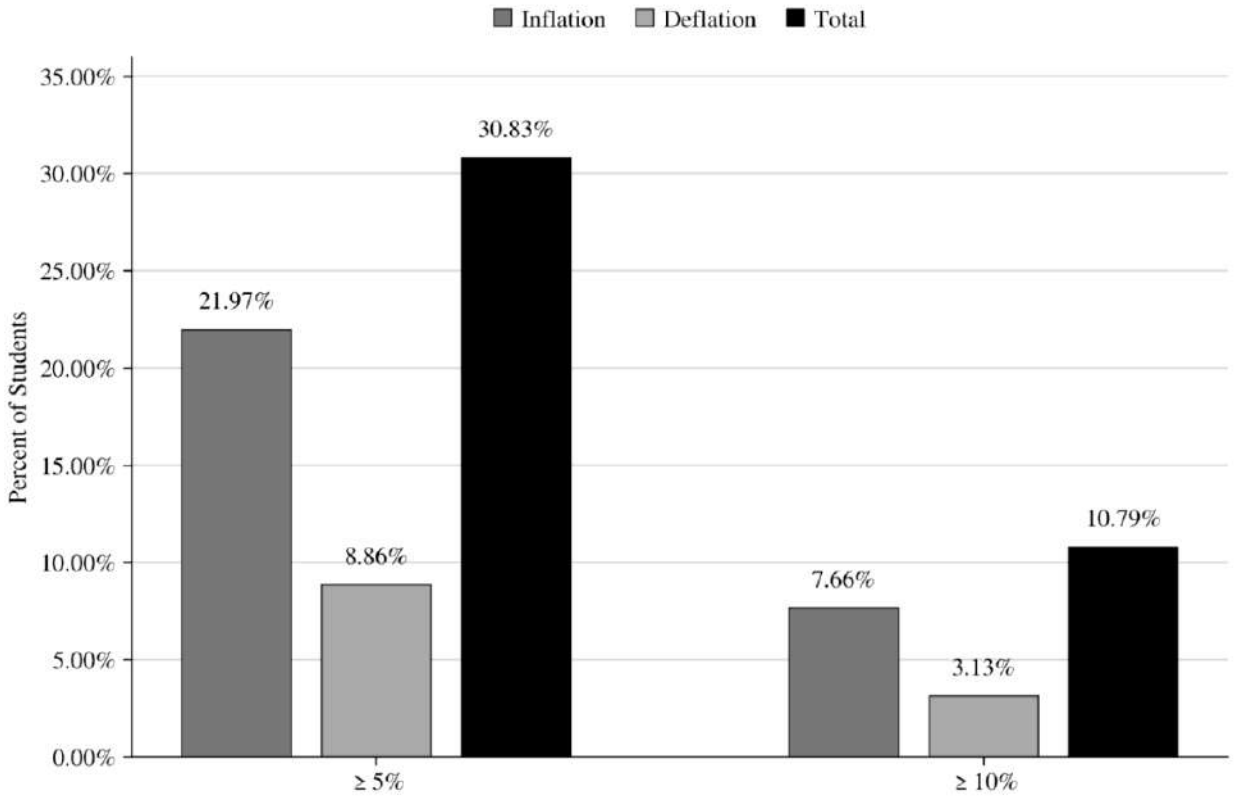
Table 3

Frequency and Percentage of Grade Inflation and Deflation

Metric	Inflation	Deflation	Total
≥ 5% Percentage Difference	21.97% (*n* = 548)	8.86% (*n* = 221)	30.83% (*n* = 769)
≥ 10% Percentage Difference	7.66% (*n* = 191)	3.13% (*n* = 78)	10.79% (*n* = 269)
≥ 1 Letter Grade Difference	28.31% (*n* = 706)	7.94% (*n* = 198)	36.25% (*n* = 904)
≥ 2 Letter Grade Difference	2.25% (*n* = 56)	0.56% (*n* = 14)	2.81% (*n* = 70)

Figure 1

Grade Discrepancies Between Final Grade Percentages and Summative Assessment Percentages

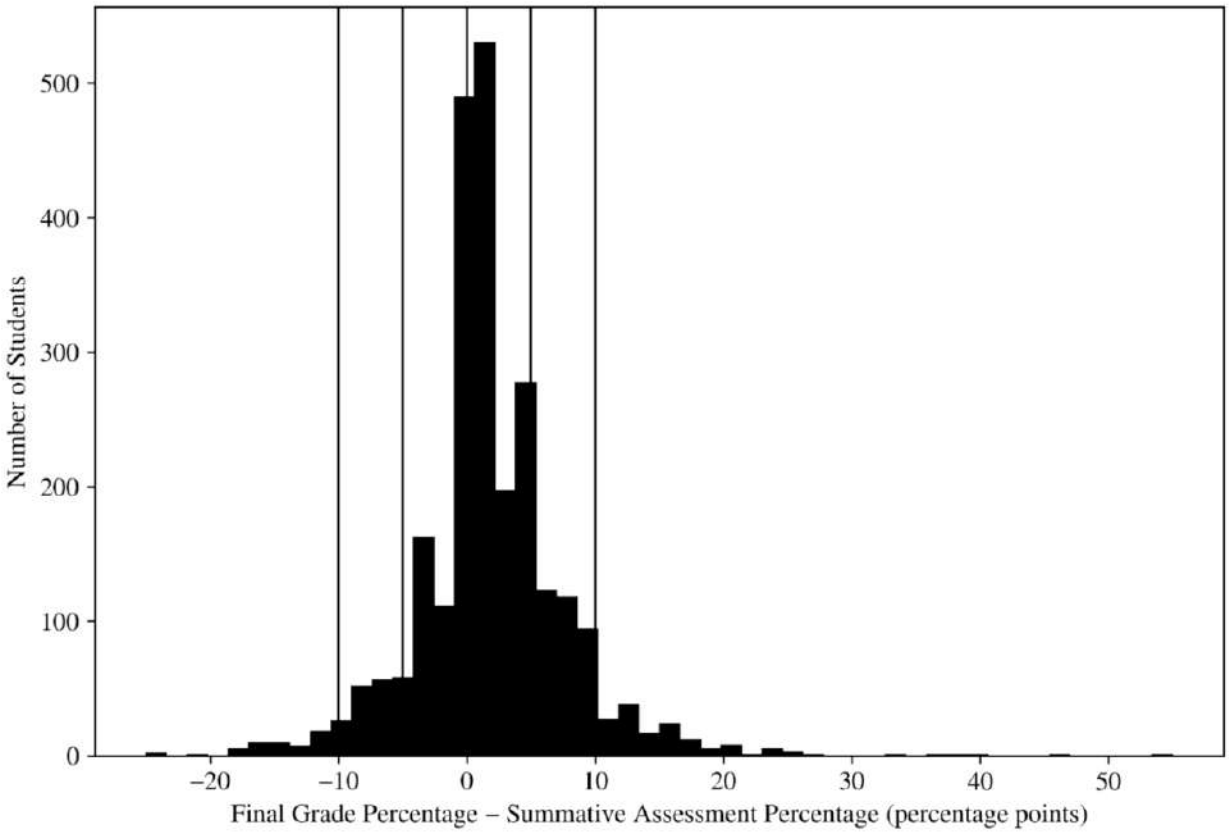


Note. Bars display percentages. *N* = 2,494. ≥ 5% inflation *n* = 548, ≥ 5% deflation *n* = 221, total *n* = 769; ≥ 10% inflation *n* = 191, ≥ 10% deflation *n* = 78, total *n* = 269.

Most discrepancies involved a one-letter grade difference; however, a smaller subset of students experienced even larger variations. Specifically, 2.81%, or 70 students, had a discrepancy of two or more letter grades. Within this group, 2.25%, or 56 students, had their final grades inflated by two letter grades. This means that a student's grade could be inflated from a C to an A. Comparatively, only 0.56%, or 14 students, experienced deflation by two letter grades, such as from a B to a D. These categorical discrepancies carry meaningful academic consequences. A one or two-letter increase could enhance a student's academic standing in transcripts and GPA, while even minor deflation may adversely affect confidence, motivation, scholarship, and other funding sources and perceptions of equity (Brookhart, 2011; Feldman, 2018; Guskey, 2020; Reeves, 2008). The disproportionate ratio of grade inflation to deflation (approximately 3.6 to 1) highlights systemic patterns and reinforces the need for greater consistency and transparency in grading practices. See Figure 3 for the grade discrepancies between final letter grades and summative assessment letter grades.

Figure 2

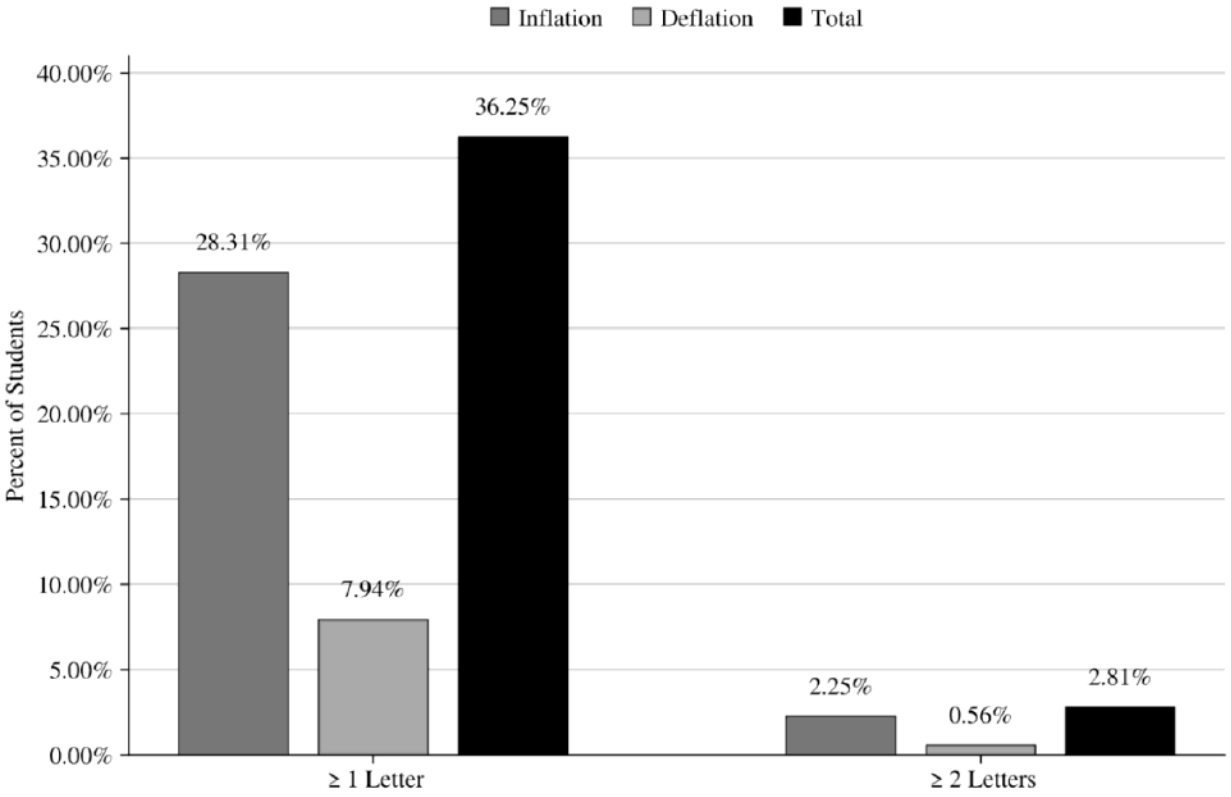
Distribution of Discrepancy Scores Across the Full Sample



Note. Discrepancy scores were calculated as final grade percentage minus summative assessment percentage (in percentage points). Positive values indicate that final grade percentages were higher than summative assessment percentages, and negative values indicate that final grade percentages were lower than summative assessment percentages. Vertical reference lines indicate −10, −5, 0, +5, and +10 percentage-point discrepancies.

Figure 3

Grade Discrepancies Between Final Letter Grades and Summative Assessment Letter Grades



Note. Bars display percentages. $N = 2{,}494$. ≥ 1 letter inflation $n = 706$, ≥ 1 letter deflation $n = 198$, total $n = 904$; ≥ 2 letters inflation $n = 56$, ≥ 2 letters deflation $n = 14$, total $n = 70$.

DISCUSSION AND IMPLICATIONS

To what extent do final grades, measured both as percentages and as letter grades, differ from summative assessment scores, indicating grade inflation or deflation? Using the analytic sample of $N = 2{,}494$, we observed an overall association between final grades and summative assessment grades but frequent and asymmetric discrepancies. At the ≥ 5% threshold, inflation occurred for 21.97% of final grades ($n = 548$; 95% CI [20.35, 23.60]) and deflation for 8.86% of final grades ($n = 221$; 95% CI [7.75, 9.98]); the combined inflation and deflation rate was 30.83% ($n = 769$; 95% CI [29.02, 32.65]). At ≥ 10%, inflation was 7.66% ($n = 191$; 95% CI [6.61, 8.70]) and deflation 3.13% ($n = 78$; 95% CI [2.44, 3.81]); the total was 10.79% ($n = 269$; 95% CI [9.57, 12.00]). For letter thresholds, at ≥ 1 of a letter grade, inflation was 28.31% ($n = 706$; 95% CI [26.54, 30.08]) and deflation 7.94% ($n = 198$; 95% CI [6.88, 9.00]); the total was 36.25% ($n = 904$; 95% CI [34.36, 38.13]). At ≥ 2 letter grades, inflation was 2.25% ($n = 56$; 95% CI [1.66, 2.83]) and deflation 0.56% ($n = 14$; 95% CI [0.27, 0.85]); the total was 2.81% ($n = 70$; 95% CI [2.16, 3.45]). This study revealed a statistically significant and practically meaningful pattern of grade inflation and deflation, confirming concerns raised in prior research about the accuracy and consistency of traditional grading practices (Brookhart et al., 2016; Griffin & Townsley, 2021; Guskey & Link, 2018).

Grade inflation was more common than deflation, consistent across both percentage and letter grade comparisons. These findings directly address the research question: final course grades often do not accurately reflect summative test results when non-academic factors are taken into account, with a substantial minority of students experiencing discrepancies in favor of higher grades. These discrepancies have essential implications for the meaning and fairness of grades. Grades are a primary currency in education, influencing high-stakes decisions such as GPA, honors, scholarships, graduation, and college admissions (Brookhart et al., 2016; Guskey & Link, 2018; Nowruzi, 2021). If final grades systematically overestimate achievement, they can provide stakeholders with an inaccurate understanding of students' true proficiency and mastery of learning. Our findings reinforce longstanding concerns in the literature that traditional grading can be inconsistent

and incorporates non-cognitive behaviors beyond academic mastery (Brookhart et al., 2016; Duncan & Noonan, 2007; Guskey, 2011; Guskey & Link, 2018; McMillan, 2001; Nowruzi, 2021; Nowruzi & Amerian, 2020).

In a century-spanning review, Brookhart et al. (2016) concluded that teacher-assigned grades are often subjective and not reliably aligned with standardized measures of learning. This study's evidence of grade inflation and deflation aligns with that conclusion, suggesting that many teachers award course grades based on a mix of achievement and non-cognitive behaviors, resulting in grades that do not strictly reflect proficiency of summative assessment performance that assesses state or national standards (Brookhart, 2011; Guskey, 2002).

Guskey and Link (2018) similarly argued that large inconsistencies in grading practices, such as what counts towards a grade and how different teachers weigh components, are a fundamental obstacle to fairness in grading. Our results support this, as the variation of grade versus summative assessment alignment indicates a lack of a uniform standard for what a course grade represents.

The prevalence of grade inflation and deflation observed here is consistent with recent empirical studies. For example, in a single-district study of high school math, Griffin and Townsley (2021) found that when homework and "employability" points were included in grades, 61.5% of students ended up with higher final grades than their test scores alone would justify, whereas about 38% saw grade deflation. They reported that 43% of students had grade differences of at least five percentage points, and 12.6% differed by a full letter grade when compared to a summative-only grade calculation. This study, which examined a broader set of courses that included all core area courses (English, math, science, and social studies), found roughly 31% with ≥ 5% discrepancies and ~36% with ≥ 1-letter discrepancies, with inflation more prevalent, a pattern similar to Griffin and Townsley's (2021) findings. This convergence suggests that the tendency for course grades to be higher than summative assessment performance is widespread and not confined to a single subject area or school. Grade inflation can mean that some students receive passing or even better course grades despite not being proficient in key content and state standards.

These outcomes exemplify why our findings matter: They highlight how current grading practices may inflate student records, potentially giving a distorted sense of student achievement or masking situations where a student is struggling with the content. To illustrate the stakes under a typical grading scale (e.g., 90/80/70), a student averaging 79% on summative assessments but receiving an 89% final (≥ 10% inflation) records a B+ rather than evidence consistent with a C, directly affecting GPA thresholds, honors lists, and scholarship eligibility. Our study strengthens the case made by prior researchers that the point-accumulation model often produces a "hodgepodge" of academic and nonacademic factors, distorting communication of what students have learned and what they are proficient in (Brookhart et al., 2016; Duncan & Noonan, 2007; Guskey, 2011; Guskey & Link, 2018; McMillan, 2001; Nowruzi, 2021; Nowruzi & Amerian, 2020). Because many gradebooks include achievement alongside non-academic factors, frequent total discrepancies of ≥ 5% and ≥ 1 letter appear even when correlations with summative evidence are high.

Taken together, these findings warrant caution in interpreting final course grades as precise measures of proficiency. Therefore, we outline key limitations and future directions.

LIMITATIONS AND FUTURE RESEARCH

While the findings of this study offer meaningful insights, it is important to acknowledge several limitations that may affect the interpretation and generalizability of the results. First, this research was conducted using gradebook data from a single school district. Because the school district's population, curricular standards, and grading culture may differ from those in other contexts, this may limit the generalizability of the findings. Beyond this, final course grades were collected from teacher-managed electronic gradebooks, introducing potential fidelity issues. We assumed that gradebook records were accurate and consistently coded; however, there may be inconsistencies in teacher practices when using gradebook software, such as data entry errors.

Although the research team used a standardized extraction protocol and cross-checking process to identify summative assessments from teacher-assigned gradebook category and assignment labels, the study examined classifications as they were operationalized in authentic classroom gradebook practice, rather than applying an external validation or reclassification process. While we acknowledge the value of examining gradebook classifications as practiced authentically in classrooms, this approach risks conflating teacher labeling inconsistencies with true assessment characteristics, potentially inflating interrater variability beyond what standardized protocols would reveal (Brimi, 2011). To strengthen validity, future studies could incorporate a brief expert review of a subsample using established reliability benchmarks. We also did not control for differences in teachers' grading scales or weighting policies beyond the summative assessments; thus, some portion of the observed discrepancies could stem from technical or recording inconsistencies rather than

intentional grading policies. Although course syllabi were not formally reviewed, the gradebook records analyzed captured teacher-reported final grade outcomes, including the effects of course-level grading practices visible in the gradebook (e.g., added points or extra credit).

It is also important to note that we used students' summative assessment averages as a benchmark for academic proficiency. This is a proxy, and it may not perfectly capture all facets of student learning or state/national standards. If a summative assessment was of uneven quality, narrow in scope, or misaligned with standards, a discrepancy between the course grade and the summative assessment may not truly indicate grade inflation. In other words, some "inflation" might reflect limitations of the assessments themselves (e.g., a student demonstrated a standard in the homework that the summative assessment did not measure). The study's design cannot disentangle true grade inflation from misalignment caused by misaligned summative assessments.

Similarly, we were unable to account for all factors that might influence grade discrepancies in a point-based grading system. Variation in grading policy among teachers could vary and not be accounted for in our data. In addition, student-level factors such as motivation or access to support might systematically affect both a student's class participation and their summative assessment performance. Without data on these variables, our analysis cannot determine which specific factors contribute most to the inconsistencies. This is a methodological limitation that leaves some ambiguity about the causes of variation in grade inflation and deflation detected in the findings.

These conclusions highlight several areas where additional inquiry is needed, prompting the following recommendations for future research. First, it remains unclear whether grading inconsistencies affect all students and subjects uniformly. Because the data spans multiple content areas analyzed in aggregate (English, math, science, and social studies), subject-specific patterns were not fully examined. Future research could investigate how grade discrepancies differ across subject areas or impact specific student groups. Additionally, employing a qualitative lens to this topic may provide a more nuanced and complex picture of the cause of grade inflation or deflation. For example, interview or observational studies might explore teacher philosophies and rationales that justify grades that diverge from summative assessment scores (e.g., What rationales do teachers give for including non-academic factors into final course grades?)

This study confirms that final course grades in point-accumulation models frequently diverge from students' demonstrated summative assessment performance, with grade inflation occurring more regularly than grade deflation. These discrepancies raise important concerns about the accuracy, fairness, and equity of point-based grading practices in K–12 education. For educational leaders, these findings suggest several actionable steps to support students, educators, and families in this work.

Educational leaders can begin by clarifying the purpose of grades and establishing shared expectations that course grades should primarily communicate academic proficiency. Leaders may also support greater grading consistency by adopting common definitions for summative and formative assessments, providing professional learning focused on separating academic achievement from behavioral compliance, and auditing gradebook practices to examine alignment between summative evidence and final grades. Leaders can promote equity by encouraging transparent grading criteria, limiting the influence of non-academic factors on grades, and using summative assessment data as a reference point when reviewing grading policies. These steps not only improve the technical accuracy of grades but also help ensure that grading practices do not unintentionally advantage or disadvantage particular groups of students. It is clear that the work behind grades needs attention (Griffin & Townsley, 2021). The findings of this study highlight the need for intentional, system-wide shifts towards evidence-based grading practices so grades can serve as valid, transparent, and equitable indicators of student learning for all stakeholders. This study can serve as a grounding framework to begin these conversations. Educators of all levels, students, and communities need to reimagine the purpose of grades, and that starts with hard conversations about the inequities already occurring.

REFERENCES

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 78*(5), 218–223. https://doi.org/10.3200/TCHS.78.5.218-223

Allensworth, E., & Clark, K. (2020). High school GPA: Predicting success in college. *Educational Researcher, 49*(6), 393–397. https://doi.org/10.3102/0013189X20902110

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. https://doi.org/10.1002/9780470743386

Brimi, H. M. (2011). Reliability of grading high school work in English. *Practical Assessment, Research & Evaluation, 16*(17). https://doi.org/10.7275/4z0x-7z

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35–36. https://doi.org/10.1111/j.1745-3992.1991.tb00182.x

Brookhart, S. M. (2004). *Grading*. Pearson/Merrill/Prentice Hall.

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership, 69*(3), 10–14.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803–848. https://doi.org/10.3102/0034654316672069

Brookhart, S. M., McTighe, J., Wiliam, D., & Guskey, T. R. (2020, September 1). Eight essential principles for improving grading. *Educational Leadership, 78*(1). https://www.ascd.org/el/articles/eight-essential-principles-for-improving-grading

Camara, W. J., Kimmel, E. W., Scheuneman, J. D., & Sawtell, E. A. (2003). *Whither the SAT? Trends in college admissions and student performance*. College Board.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.

Duncan, R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *The Alberta Journal of Educational Research*, *53*(1), 1–21. https://doi.org/10.11575/ajer.v53i1.55195

Feldman, J. (2018). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin Press.

Galla, B. M., Shulman, E. P., Plummer, B. D., Gardner, M., Hutt, S. J., Goyer, J. P., D'Mello, S. K., Finn, A. S., & Duckworth, A. L. (2019). Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability. *American Educational Research Journal, 56*(6), 2077–2115. https://doi.org/10.3102/0002831219843292

Gershenson, S. (2018). *Grade inflation in high schools (2005–2016)*. Thomas B. Fordham Institute.

Godfrey, K. E. (2011). *Investigating grade inflation and non-equivalence* (College Board Research Report No. 2011-2). College Board. https://eric.ed.gov/?id=ED521179

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *The Journal of Experimental Education, 74*(3), 249-266. https://doi.org/10.3200/JEXE.74.3.249-266

Griffin, R. T. (2020). *Grading and equity: Inflation/deflation based on race, gender, socio-economic and disability statuses when homework and employability scores are included* (Doctoral dissertation, University of Northern Iowa). UNI ScholarWorks. https://scholarworks.uni.edu/etd/1045

Griffin, R., & Townsley, M. (2021). Points, points, and more points: High school grade inflation and deflation when homework and employability scores are included. *Journal of School Administration Research and Development, 6*(1), 1-11. https://doi.org/10.32674/jsard.v6i1.3460

Guskey, T.R. (1994). Making the grade: What benefits students? *Educational Leadership, 52*(2), 14-20.

Guskey, T. R. (2002). *Developing grading and reporting systems for student learning*. Corwin Press.

Guskey, T. R. (2009). *Practical solutions for serious problems in standards-based grading*. Corwin Press.

Guskey, T. R. (2011). *Five obstacles to grading reform*. *Educational Leadership, 69*(3), 16–21. https://www.ascd.org/el/articles/five-obstacles-to-grading-reform

Guskey, T. R. (2013). The case against percentage grades. *Educational Leadership, 71*(1), 68–72.

Guskey, T. R., & Link, L. J. (2018). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice, 26*(3), 303–320. https://doi.org/10.1080/0969594X.2018.1555515

Guskey, T. R. (2020). Breaking up the grade. *Educational Leadership, 78*(1), 40–46. https://www.ascd.org/el/articles/breaking-up-the-grade

Knight, M., & Cooper, B. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment and student behavior. *NASSP Bulletin, 103*(1), 3–21. https://doi.org/10.1177/0192636519826709

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Link, L. J., & Guskey, T. R. (2022). Is standards-based grading effective? *Theory into Practice, 61(*4), 406–417. https://doi.org/10.1080/00405841.2022.2107338

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Prentice Hall.

Marzano, R. J., & Heflebower, T. (2011). Grades that show what students know. *Educational Leadership, 69*(3), 34–39.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20–32. https://doi.org/10.1111/j.1745-3992.2001.tb00055.x

McMillan, J. H., & Hellsten, L. (2010). *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson Education Canada.

McMillan, J. H. (2018). *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson.

Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan, 96*(7), 64–68. https://doi.org/10.1177/0031721715579043

Nowruzi, M., & Amerian, M. (2020). Exploring the factors Iranian EFL institute teachers consider in grading using personal construct theory. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills), 38*(4), 123–164. https://doi.org/10.22099/jtls.2020.36293.2780

Nowruzi, M. (2021). A study of EFL teachers' classroom grading practices in secondary schools and private institutes: A mixed methods approach. *Language Testing in Asia, 11*, Article 29. https://doi.org/10.1186/s40468-021-00145-2

O'Connor, K. (2009). *How to grade for learning, K–12* (3rd ed.). Corwin.

Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin, 78*(561), 55–64. https://doi.org/10.1177/019263659407856111

Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher, 42*(5), 259–265. https://doi.org/10.3102/0013189X13481382

Pollio, M., & Hochbein, C. (2015). The association between standards-based grading and standardized test scores as an element of a high school reform model. *Teachers College Record, 117*(11), 1–28. https://doi.org/10.1177/016146811511701106

Reeves, D. B. (2008). Leading to change: Effective grading practices. *Educational Leadership, 65*(5), 85–87.

Reibel, A. R., Twadell, E., Onuscheck, M., & Gobble, T. (2025). *Pathways to proficiency: Implementing evidence-based grading* (2nd ed.). Solution Tree.

Schinske, J., & Tanner, K. (2017). Teaching more by grading less (or differently). *CBE—Life Sciences Education, 13*(2), 159–166. https://doi.org/10.1187/cbe.cbe-14-03-0054

Schneider, J., & Hutt, E. (2014). Making the grade: A history of the A–F marking scheme. *Journal of Curriculum Studies, 46*(2), 201–224. https://doi.org/10.1080/00220272.2013.790480

Starch, D., & Elliott, E. C. (1912). Reliability of grading high school work in English. *School Review, 20*(7), 442–457.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education, 4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Townsley, M. (2018). Mastery-minded grading in secondary schools. *School Administrator, 75*(1), 16–19.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wilkinson, L., & the APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594

Woodruff, D. J., & Ziomek, R. L. (2004, March). *High school grade inflation from 1991 to 2003* (ACT Research Report Series 2004–04). ACT, Inc. https://files.eric.ed.gov/fulltext/ED484784.pdf

Wormeli, R. (2017). *Fair Isn't Always Equal: Assessment & Grading in the Differentiated Classroom* (2nd ed.). Routledge. https://doi.org/10.4324/9781032681122

Yeritsyan, A., Mjelde, J. W., & Litzenberg, K. K. (2022). Grade inflation or grade increase. *Journal of Agricultural and Applied Economics, 54*(2), 375–393. https://doi.org/10.1017/aae.2022.15

ALLIE KOLLMAN, MEd, is an educational consultant, supporting educators across grade levels in the development and implementation of instructional practices and district-wide initiatives. Her research interests include evidence-based grading, grading reform, assessment practices, and science education. She brings a background in teaching middle and high school biological and physical sciences, along with experience as a laboratory assistant and researcher in the biological sciences. Email: alliekollman@gmail.com

ERIN SINNER, MEd, is a high school assistant principal in North Dakota and a doctoral candidate in educational leadership at North Dakota State University. Her research interests center on grading reform, instructional equity, and the implementation of evidence-based grading practices in secondary schools. Before her current role, she served as a high school instructional coach and English teacher. Erin is passionate about making grading practices more accurate, fair, and student-centered. Email: sinnere@fargo.k12.nd.us

HANNAH NURMI, is an instructional coach in North Dakota and a doctoral candidate in educational leadership at North Dakota State University. Her work centers on advancing systemic change to ensure all students, regardless of background, have access to rigorous academic opportunities. As an instructional coach, she believes that reflective dialogue and intentional support empower teachers to challenge inequitable practices and cultivate more just, student-centered classrooms. Prior to her coaching role, Hannah taught high school English. Email: nurmih1@fargo.k12.nd.us

CAILEN O'SHEA, PhD, is an assistant professor of educational and organizational leadership at North Dakota State University. His research interests focus on school transformation, specifically enhancing instruction to provide more equitable settings for all students. Previously, Dr. O'Shea served as a behavior interventionist, 5th-grade teacher, and instructional technology coach in Title I schools in Lincoln, Nebraska. Email: cailen.oshea@ndsu.edu

Manuscript submitted: ***September 7, 2025***
Manuscript revised: ***March 1, 2026***
Accepted for publication: ***March 15, 2026***

***Special Issue**: Exploring Current & Future Grading Practices in K-12 Schools*
Best Practice Article

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 121-130
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

Boolean Logic Gates: A New Method of Grade Calculation for Standards-Based Grading

Shawn Carney
Fargo Public Schools, USA
Teresa Shume
North Dakota State University, USA

ABSTRACT

Standards-based grading (SBG) has emerged as a reform movement aimed at improving the accuracy, equity, and transparency of grading systems. SBG involves grading at two levels: the overall course or subject area, and individual standards. This article examines multiple methods for calculating grades for individual standards in SBG systems, including logic rules, reliance on most recent evidence, various types of means, power law projections, and frequency-based approaches such as mode and bi-modal matrices. Each method is analyzed based on four criteria: quantity of assessment evidence, recency of assessment evidence, how easily and transparently students can track their own progress, and whether students who learn content more slowly are disadvantaged by the grading method. A novel method, Boolean logic gates, is proposed for determining a student's level of proficiency on an individual standard. By systematically incorporating conditional logic, Boolean logic gates are a promising and innovative grade calculation method that has the potential to provide educators with nuanced and actionable representations of student learning.

***Keywords*:** grading, standards-based grading, classroom assessment

Many schools today are seeking to reform their grading systems by shifting away from traditional grading and implementing standards-based grading (SBG) (Peters & Buckmiller, 2014). The central purpose of grades is to communicate the extent to which students have mastered the learning targets; thus, accuracy is a key tenet that underpins effective grading systems (Brookhart et al., 2016). Traditional grades are typically based on calculating the mean of an accumulation of points earned through various assessment methods (e.g., homework, quizzes, tests, projects), as well as compliance with various classroom policies and expectations (Guskey & Jung, 2012). Traditional grading has been criticized as invalid and unreliable for measuring and reporting students' academic achievement, and has been called "a hodgepodge of attitude, effort, and achievement" (Brookhart, 1991, p. 36). SBG, in contrast, reflects students' understanding based on the standards being taught. SBG approaches are not obscured by non-academic factors such as class participation, effort, or meeting deadlines (Townsley, 2019). Moreover, homework and other formative assessments intended to support students on their pathway of learning are also excluded from SBG (Brookhart et al., 2020). Consequently, SBG systems provide a more accurate, equitable, and actionable representation of what a student knows (Guskey et al., 2024).

While a variety of approaches to SBG exist (Link & Guskey, 2022), Townsley and Wilcox (2024) have identified four core principles that distinguish SBG from traditional grading approaches. For the purposes of this article, SBG will be defined as grading systems that adhere to the following key tenets provided by Townsley and Wilcox (2024, p. 76):

1. Grading should communicate students' current level of learning based on standards.
2. Teachers should use a finite number of performance categories to assess students' learning.
3. Grades should be based on summative assessments rather than formative assessments or student behaviors.

4. Students should have multiple opportunities to demonstrate their learning.

The definition from Townsley and Wilcox (2024) was chosen for this article because it is current and mostly agreed upon by SBG scholars.

Grading at Two Levels

As school leaders navigate the challenges of establishing SBG systems, crucial decisions must be made to select specific methods for calculating standards-based grades (Heflebower et al., 2018). These decisions typically occur at two levels: conveying students' overall learning in a course or subject area and communicating a summary of student proficiency for each standard. At the overall course level, an informative way to communicate student learning is to present proficiency statements for each individual standard, providing stakeholders with clear messages about what a student knows, understands, and is able to do (Iamarino, 2014). However, schools frequently encounter a pervasive need to report out course grades as single letters due to requirements for college admissions, criteria for scholarship applications, and eligibility for co-curricular activities, among other reasons (O'Connor, 2018). To translate a student's scores on individual standards into a single course letter grade, the use of "logic rules" is regarded as an appropriate and suitable method (O'Connor, 2018).

When using logic rules to determine a course grade, each conditional statement, or "rule," is applied in turn until one is found that corresponds to the student's collection of standard-level grades. Table 1 provides an illustrative example of a set of logic rules that could be applied to determine a course grade. In this example, "4" indicates advanced, "3" indicates proficient, "2" indicates partially proficient, and "1" indicates developing.

Table 1

Logic Rules Example

Course Grade	Logic Rule
A	All standard scores are at the 3 or 4 level
B	One standard score of 2, with no standard scores of 1
C	More than 1 standard score of 2, with no standard score of 1
D	One standard score of 1
F	More than 1 standard score of 1

Methods to determine overall course grades, such as logic rules, are predicated on determining students' scores for each standard, the second major type of decision educators must make when implementing SBG in their school.

The aim of this article is to analyze a variety of grade calculation methods for translating a set of assessment scores pertaining to a particular standard into a single summary score for that standard. The grade calculation methods analyzed in this article include logic rules, most recent score, three types of means (calculated average, average of the three highest scores, and decaying average), the Power Law, and two types of frequencies (mode and bi-modal matrix). Finally, a novel system that uses Boolean logic gates is proposed and analyzed in comparison with the other calculation methods.

Purely professional decision-making based on evidence has been omitted because this article focuses on comparing calculation methods. Professional decision-making, simply put, is not a calculation method. It requires evaluating qualitative properties of assessments that earned quantitative scores, which cannot be numerically represented. Some SBG experts emphasize the value and importance of teachers' professional judgment in determining SBG grades (Guskey & Jung, 2016; Townsley & Wear, 2020). This article introduces a novel calculation method that incorporates professional judgment; however, it does not evaluate professional judgment as a standalone alternative to the calculation methods considered.

APPROACH TO ANALYSIS

Criteria

In this article, we analyzed a variety of grade calculation methods using four criteria derived from research literature and the authors' combined teaching experience of over four decades, consisting of 17 years for one author and 32 for the other. First, we considered the quantity of evidence used to determine the summary grade for each standard because this impacts the grade's validity. Using a single piece of evidence to determine a grade presumes that the assessment used is

entirely valid and reliable. While those are attributes to which educators should aspire, it is not feasible to completely avoid threats to validity and reliability in the design of each assessment (O'Connor, 2018). Some SBG scholars highlight the value of collecting a body of evidence to use when determining grades (e.g., Heflebower et al., 2014). However, such bodies of evidence sometimes conflate formative and summative assessments, which is problematic since SBG should be based exclusively on summative assessments (Townsley & Wilcox, 2024). In sum, using more evidence is better than using less evidence, but the quality and salience of the evidence must also be considered.

Second, we examined whether the most recent data were prioritized. Compared to initial attempts, a student's most recent scores are more likely to accurately indicate the level of proficiency attained by the student (Townsley & Wear, 2020). Some scholars recommend that only the most recent performance that occurred at the end of the grading period should be considered (Feldman, 2019). Recentness of the evidence does contribute to its salience. However, there exists a tension between recentness and quantity since relying exclusively on a single piece of evidence, even if most recent, presents limitations described above. Moreover, restricting grade calculation to the single most recent assessment creates a "high-stakes" assessment that can induce anxiety among students (Schimmer, 2016).

Third, we checked if standard summary scores produced through each calculation method incentivize student growth. Fundamental tenets of effective SBG systems include fostering student self-efficacy, contributing to student confidence, and nurturing a growth mindset (Schimmer, 2016). Beyond opportunities to practice multiple formative assessments and to repeat summative assessments, growth is incentivized by transparency when students can easily track their own progress by calculating their ongoing grade as they earn additional summative scores (O'Connor, 2018). Grading practices should convey the message that eventual success is possible, incentivizing rather than penalizing students for growth (Dueck, 2011).

Last, we considered whether students who take different amounts of time to learn the same content are awarded the same summary score for a standard. Students who eventually attain the same level of proficiency should earn the same grade, regardless of whether one student learned the content faster or slower than another (Townsley & Wear, 2020). Students' rates of learning can be influenced by numerous factors, such as learning disabilities, interruptions to formal schooling, and English language learning needs, among many others. Simply put, this is a matter of equity (Feldman, 2023).

GRADE CALCULATION METHODS

What follows is an explanation of each calculation method identified for determining a summary grade for a single standard and an analysis of each method to identify strengths and weaknesses. For added clarification, each method is applied to the following set of example student scores earned on an SBG grade scale in chronological order: 1, 1, 2, 2, 3, each representing a score for a different summative assessment of relatively comparable length and rigor pertaining to the same standard. Depending on specific instructional contexts, these scores could represent assessments of closely related applications of a single standard or a cluster of distinct subskills related to the standard. In this example, 4 indicates advanced, 3 indicates proficient, 2 indicates partially proficient, and 1 indicates developing.

Logic Rules

As described earlier, logic rules aim to identify the most consistent level of achievement through application of a series of conditional statements. For the purpose of example, one way the logic rules listed in Table 1 could be adjusted to produce a summary grade for an individual standard could be as follows: a standard-level score of 3 requires no 1s or 2s, a 2 requires one score of 2 and no scores of 1, and a score of 1 requires any amount of 1s. This rule set is quickly observed as taking the worst score from the data set and assigning it for the standard score. For argument's sake, a more lenient set of rules could be applied: a standard-level score of 3 requiring no more than one 2 and no 1s, a 2 requiring two or more 2s and no more than one 1, and a 1 requiring more than one 1. Using either set of logic rules, our student with hypothetical scores of 1, 1, 2, 2, 3 would earn a 1 as a summary score for the standard, as they have more than one score evaluated at a 1, even though their most recent score was a 3. Beyond these two examples, logic rules can be operationalized in a wide variety of ways.

The transparency of logic rules can be impacted by the timing of when the number of scores that will be used to calculate the grade is determined. Transparency of logic rules is enhanced when the number of assessment scores is set at the beginning of the marking period, yet this can be constricting to teachers when designing instruction and assessment in response to evolving student needs. Logic rules do not evaluate grades based on recency unless recency is explicitly considered; for example, a set of logic rules could be applied to only the five most recent scores. Additionally, when applying logic rules, the quantity of scores below proficient can be more impactful than the quantity of proficient scores. In effect, this causes all assessments to be "high stakes" and penalizes students who take longer to demonstrate proficiency in a standard. Further, a logic rule approach means that growth may not be incentivized, as once a student earns a lower score,

depending on the rules presented, they may be locked out of ever attaining a standard-level score of proficient. In our examples presented above, after receiving one or two scores of 1, the student is not incentivized to grow, as they will never be eligible to earn a 3 in this standard.

Most Recent Score

In this straightforward calculation, only the most recent piece of evidence is used to determine the grade. Using our sample score of 1, 1, 2, 2, 3, the student's grade for this standard would be 3 (proficient) because it was their final assessment score.

A number of SBG experts recommend the use of a single score based on the most recent evidence (Feldman, 2019). Simply taking the most recent evidence to be a student's grade, however, relies heavily on the validity of a single assessment. As Marzano (2010, p. 29) points out, "Even if a very good final assessment is designed, it will still contain error. Consequently, it cannot be absolutely trusted as the true score for an individual student." While this method prioritizes recent evidence and avoids penalizing students for learning concepts more slowly, it does not address the value of multiple assessment results forming a body of evidence (Heflebower et al., 2014). Moreover, some students may put less effort into initial assessments, knowing that scores can be later replaced. Additionally, it is possible for students to be awarded a grade that may, in fact, be an exception.

Types of Means

Reflecting a common understanding of the term "average," the calculated mean is determined by taking the sum of the scores for the pieces of evidence and dividing it by the number of pieces of evidence. Using our example scores of 1, 1, 2, 2, 3, the calculation becomes (1+1+2+2+3)/5, yielding a score of 1.8. A second approach to calculating a mean is to truncate the data set to only the three pieces of evidence with the highest scores. For the data set 1, 1, 2, 2, 3, only the 2's and the 3 would determine the student's standard-level grade. This would calculate as (2+2+3)/3 = 2.3. A third type of mean is a decaying average, where scores are weighted progressively, with the most recent score assigned the highest weight. For this example, the most recent evidence is allocated 35% of the total grade, scaling down to 22%, 20%, 13%, and finally only 10% for the first piece of evidence the student provided. For the set of scores 1, 1, 2, 2, 3 earned in this order, a student's grade calculation would appear as (1*10) + (1*13) + (2*20) + (2*22) + (3*35)/100 = 2.12.

Compared to the most recent method, all three of these methods rely on a more robust body of evidence to support a student grade, though only the decaying average prioritizes recent evidence. Regardless of whether an average is weighted or not, however, using a mean calculation penalizes students who do not achieve proficiency on their first assessment and must demonstrate proficiency beyond the proficient level in order to reach a standard-level grade of proficient, which may discourage students (O'Connor, 2018). As Schimmer (2016, p. 61) points out, "The mean can be mean!" Consequently, no method that relies on calculating any type of average can accurately represent a student's standard-level grade at the end of a grading period because of the need for higher scores to compensate for any that fall below the proficient level (O'Connor, 2018). Moreover, means that produce overall scores with decimals echo the illusion of precision perpetuated by traditional grading methods (Guskey, 2024).

Power Law

Identified as one of the main approaches to standard-level grade calculation in SBG literature (Townsley & Wilcox, 2024), the power law approach is a statistical technique that models score changes over time (Marzano, 2006). This approach is not designed to average grades, but rather to predict the next grade a student would earn based on a statistical model. Applying a power law to the data set of 1, 1, 2, 2, 3, the student grade would be projected as 2.57.

The power law approach draws on the full body of evidence available and largely relies on the sequence in which scores are earned. Because such an approach attempts to fit a data set to a linear projection, it creates a near opposite flaw compared to means. Where means can create a deficit score that penalizes students who learn material more slowly, the power law instead penalizes those who demonstrate proficiency early on. Having low scores early on and higher scores later in the grading period will lead to a higher projected grade by adjusting the slope of the projection. This can lead students to intentionally perform below proficient early on in a grading period because it will increase the projected score later in that grading period. Additionally, because the power law is a projection, a student does not necessarily need to ever achieve proficiency, only to show growth that would project them on a course to demonstrate proficiency. Finally, because of the

complexity of the power law statistical calculations, this method may not be as transparent or appealing as more easily explained grading practices (Townsley & Wilcox, 2024).

Types of Frequencies

When the mode is used to determine a grade, whichever score a student earns most frequently is the grade they are assigned. In the case of ties, a decision needs to be made to either select the higher mode, lower mode, or calculate a mean average of the two. If the option of calculating a mean average of the two is chosen, all the drawbacks of mean calculations described above permeate this system as well. Consequently, selecting the lower or higher mode should be considered instead. In the example score set of 1, 1, 2, 2, 3, the modes of 1 and 2 are tied.

A second type of mode is the bi-modal matrix, which references both the most common score and the second most common score that a student earns (Reibel et al., 2025). To determine a grade, the primary and secondary modes are determined, and then a matrix table is used to produce a final score (Reibel et al., 2025). Under each grade in the illustrative example in Table 2, corresponding primary and secondary modes appear in order. For example, a grade of 4 is assigned when a student's modes are 4 and 3, 4 and 2, or 3 and 4. For the sample score set of 1, 1, 2, 2, 3, the modes identified are 1 and 2. Deferring to the lower mode, the summary grade for the standard would be 1. Deferring to the higher mode, the summary grade for the standard would be 2.

Table 2

Example of a Bi-modal Matrix

Grade: 4	Grade: 3	Grade: 2	Grade: 1
4-3	4-1	2-3	1-4
4-2	3-2	2-4	1-3
3-4	3-1	2-1	1-2

These two grading approaches do not penalize students whose learning progress takes more time. However, both are limited by the fact that neither evaluates the entire body of evidence provided by a student nor prioritizes the most recent evidence. Further, it is possible for students to be penalized for earning higher grades that lie outside the modal value. For example, our student who earns the scores 1, 1, 2, 2, 3 over a series of assessments will still earn a 1 or 2, even if they were to score a 4 on their final assessment. For the bi-modal matrix, our student would earn a final grade of 1 if rounding down for ties between modal scores. Notably, had the student scored 2 instead of 3 on the final assessment, the primary mode would have been 2, and the secondary mode would have been 1, resulting in a summary score of 2. Ironically, this student's final grade would be higher if they had scored lower on the last assessment. A student may be disincentivized when they realize they have earned so many similar scores that their last assessments will not impact their final grade or when it is more strategic to earn a lower score.

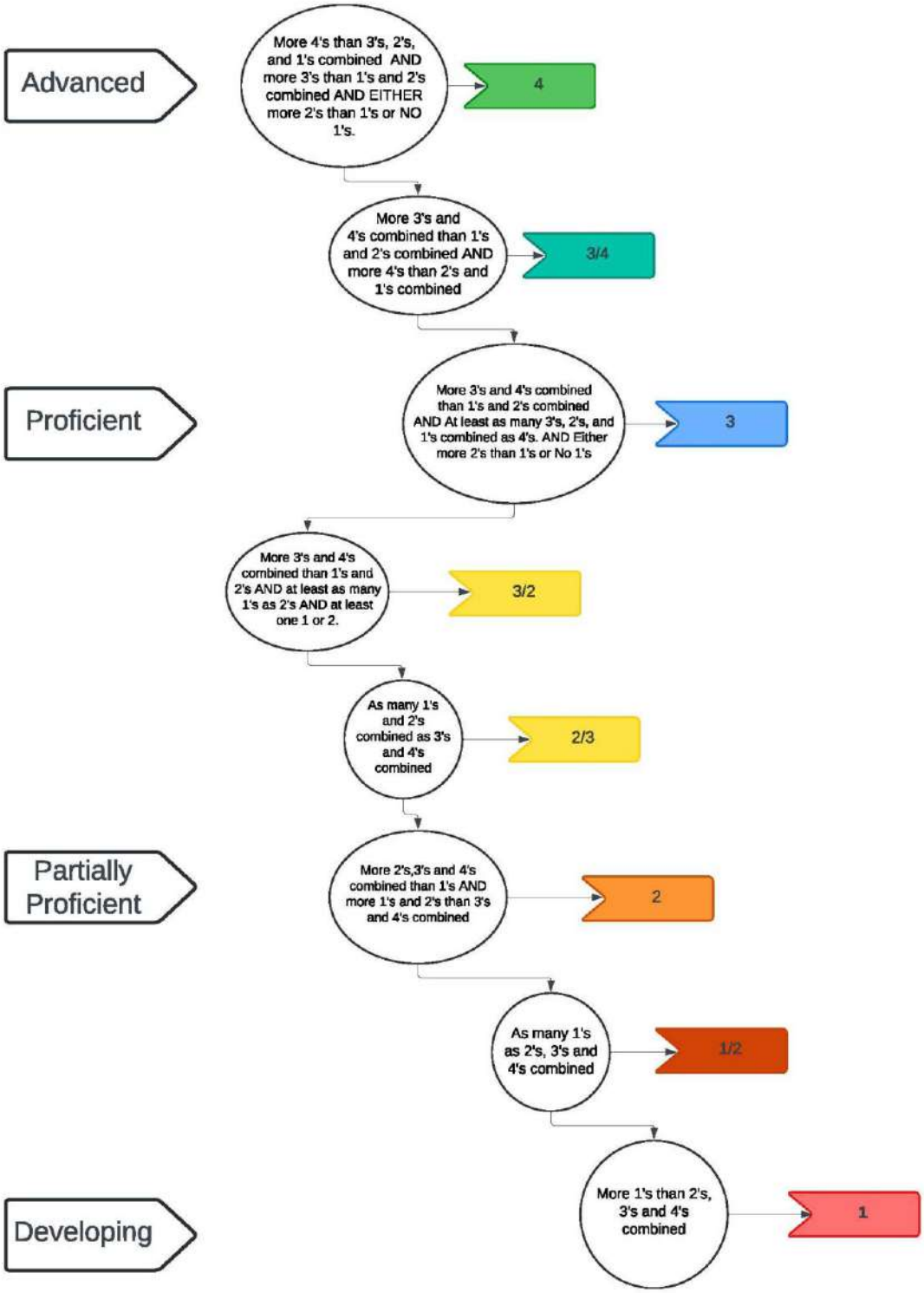
Boolean Logic Gates

Presented below is a novel approach to grade calculation for a single standard based on if-then Boolean logic gates. To interpret a set of scores based on Boolean logic gates, refer to the example provided in Figure 1. If a statement is found to be true, the resulting grade appears in a flag to the right. If the statement is found to be false, progress to the next bubble until a true statement is located. Notably, the Boolean logic gate system identifies instances where the student's grade appears as a split score. In these situations, the teacher should use professional judgment to determine which score best represents the student's proficiency level. Figure 1 offers one example of a Boolean logic gate system; other variants are possible.

Utilizing true/false statements in a similar manner to logic rules used to determine a course grade, Boolean logic gates chain together multiple Boolean statements to analyze a full body of evidence of student learning to determine a grade for a single standard. Compared to logic rules, Boolean logic gates use comparative operators rather than absolutes to evaluate the capacity and frequency of student demonstrations of proficiency. Unlike logic rules, which typically operate purely on the basis of the presence of low scores, Boolean logic gates account for the frequency of higher scores. This allows for bodies of evidence wherein demonstrations of proficiency can have varied levels of success and permits adjustments to the

Figure 1.

Example of a Boolean Logic Gates Grade Calculation



quantity of assessments during a grading period. Additionally, designing grade calculation as an ordered series of gates, rather than a table of separate rules, guides stakeholders to more easily understand the order of progression and what is required to move from one proficiency level to the next, thus promoting greater transparency.

In Boolean logic gate systems, a student's final grade accounts for all data points collected and weighs them against lower-level demonstrations, creating a score that uses more data than modes and avoids the penalizations of means. As pointed out previously, means can create insurmountable obstacles for students to recover from if poor levels of proficiency are demonstrated early on. Boolean logic gates do not require students to show demonstrations above proficient as a mean would in order to account for lower scores; rather, they only require students to demonstrate proficient more often than not. Additionally, modes fail to accurately represent student proficiency because they do not represent the full body of evidence. When a variety of scores are earned in a modal calculation, the least frequent scores become lost, which penalizes students if those scores were above their modal score.

Notably, some variants of Boolean logic gate systems can reveal student growth earlier than other methods. Especially in cases where students have wider distributions of scores, a student who struggles early on may not show even partial proficiency until late in a grading period. For example, our example student who chronologically earns a score set of 1, 1, 2, 2, 3, for a standard will never see their grade projected as higher than 1 if calculated using a modal or bi-modal method. This can be harmful to students who are improving yet do not see their projected grade increase. Within the Boolean logic gate example depicted in Figure 1, this student starts with the same scores, but will see their grade projected as a 1-2 split after the 4th assessment, and 2 after the fifth, encouraging growth and engagement. Finally, unlike all other methods analyzed, Boolean logic gates identify cases with split scores where professional judgment should be applied. Incorporating professional judgment into a calculation method allows for recognition that scores are not simply numbers, but rather are representations of learning (O'Connor, 2018).

As with any grading system, there are trade-offs to consider. For example, like logic rules, Boolean logic gates do not recognize recency of scores. Also, some variants of Boolean logic gates, such as the one depicted in Figure 1, can result in a specific scenario where, regardless of other scores, a student who has more 1s than 2s cannot earn a rating of proficient or advanced. A learner in such a situation could feel discouraged, and student growth might not be incentivized.

COMPARING APPROACHES

To more fully distinguish between the methods analyzed in this article, scores for four sample students are calculated in Table 3 below. Each student earned seven scores pertaining to a particular standard. Like the previous set of sample student scores, it can be assumed these are summative assessments of relatively comparable length and rigor and are listed in chronological order. Also, these scores are based on the same four-level proficiency scale.

Students C and D highlight the value of considering recency in grades. While both students earned the same number of scores of each level of proficiency, they earned those scores in different orders. Student D shows growth while Student C shows that they understood content early on, but their ability to demonstrate a standard diminished overtime. Some methods, including Boolean logic gates, did not pick up on this distinction.

Students A and B highlight the value of using all data in grades. While both students demonstrate the same primary and secondary modal scores as well as logic rules scores, the remaining scores show they have quantitatively different understandings of course material. Additionally, these two students demonstrate that recency on its own may not be enough to evaluate a student's level of understanding. Because all assessment scores include potential error, a singular grade may be fallible (Marzano, 2010). In a course where competencies or skills within a standard do not vary greatly, the use of more recent evidence may be helpful. However, in a course where each of these pieces of evidence is a demonstration of content of varied difficulties, recency may not provide a sufficient picture.

Table 3

Comparison of Grade Calculation Methods for a Single Standard

Grading Method	Student A (4,3,2,2,4,3,2)	Student B (2,2,2,1,3,3,4)	Student C (3,2,3,1,1,2,1)	Student D (1,1,3,2,1,2,3)
Logic Rule (original)	2	2	1	1
Logic Rule (lenient)	2	2	1	1
Most Recent	2	4	1	3
Mean	2.9	2.4	1.9	1.9
Mean of Highest 3	3.7	3.3	2.6	2.6
Decaying Average	2.7	3.0	1.5	2.2
Power Law	2.4	2.8	1.2	2.3
Mode	2	2	1	1
Bi-modal Matrix	2	2	1	1
Boolean Logic Gates	3-4	2	2	2

Note. Calculations adhere to the methods described in each corresponding section. When a calculation resulted in a decimal value, the value was rounded off to the nearest tenth.

LIMITATIONS

The Boolean logic gate method is a novel conceptual idea that has not been empirically tested. Pilot testing and research studies need to be designed and implemented before widespread adoption could become feasible. Additionally, like many other SBG calculation methods, the Boolean logic gate method may not be compatible with typical student information systems and electronic gradebooks commonly used in schools today. Further, before widespread implementation could be possible, consideration would need to be given to teacher training, professional learning community calibration, and communicating this calculation method to students and families. Because Boolean logic gate systems can appear complex at first glance, it would be important to ensure all stakeholders understand the model.

CONCLUSION

As schools navigate the process of SBG grading reform, school-based educators need to select grade calculation methods to determine not only overall course grades, but also summary scores on each standard. Trusting the most recent score as the singular piece of evidence relies entirely on the validity of a single assessment. The power law approach generates a projection that diminishes transparency for stakeholders because of its complexity (Townsley & Wilcox, 2024). Means can penalize students for low scores (O'Connor, 2018) and risk perpetuating the illusion of precision perpetuated by traditional grading systems (Guskey, 2024). Approaches based on modes leave out evidence and can discourage students if they determine the remaining assessments will have little effect on their overall grade or when earning a lower grade becomes the more strategic choice.

The newly proposed Boolean logic gates method, on the other hand, draws on advantages of means and modes while avoiding their pitfalls. Like mean calculations, Boolean logic gate grades account for all available evidence. Like mode calculations, Boolean logic gate grades do not provide false precision of decimal values and do not require that students score higher than proficient in order to account for low scores earned early in their learning. Means show growth earlier by having every score contribute to the ongoing grade. Modes, in turn, can be narrow in scope, requiring significant amounts of data to reflect progress in student learning. The Boolean logic gate method strikes a balance between these, showing growth that may incentivize students earlier than modal calculations without trying to interpolate a value between levels of proficiency, as mean calculations do. For these reasons, the Boolean logic gates grade calculation method for individual standards is poised to offer an informative representation of student learning and shows strong potential as a useful tool for school-based educators implementing SBG systems.

School leaders who wish to explore the possibility of adopting Boolean logic gates into their school or district's grading practices could begin by leading a gradebook investigation. Such an undertaking would involve applying a variety of grading methods, including Boolean logic gates, to samples of prior gradebooks and leading a discussion with teachers to examine hypothetical impacts on student grades. Guiding teachers to examine prior gradebooks through the lenses of various grade calculation methods can evoke deeper discussions about the meaning of grades. Schools have complex decisions to make about grade calculation methods. Boolean logic gates, like any grade calculation method, introduce strengths and limitations. School and district administrators can provide valuable leadership when working collaboratively with educators to make context-specific decisions about the trade-offs that best align with schools' needs and priorities.

REFERENCES

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35–36. https://doi.org/10.1111/j.1745-3992.1991.tb00182.x

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803–848. https://doi.org/10.3102/0034654316672069

Brookhart, S. M., Guskey, T. R., McTighe, J., Wiliam, D. (2020). Eight essential principles for improving grades. *Educational Leadership, 78*(1), 10–17.

Dueck, M. (2011). How I broke my rule and learned to give retests. *Educational Leadership, 69*(3), 72–75.

Feldman, J. (2019). Beyond standards-based grading: Why equity must be part of grading reform. *Phi Delta Kappan*, *100*(8), 52-55. https://doi.org/10.1177/0031721719846890

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

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

Guskey, T. R., Frey, N., & Fisher, D. (2024). *Grading with integrity: A research-based approach grounded in honesty, transparency, accuracy, and equity.* Corwin.

Guskey, T. R., & Jung, L. A. (2012). The case against averaging grades. *Principal Leadership, 13*(6), 28–32.

Guskey, T. R., & Jung, L. A. (2016). Grading: Why you should trust your judgment. *Educational Leadership, 73*(7), 50–54.

Heflebower, T., Hoegh, J., & Warrick, P. (2014). *A school leader's guide to standards-based grading*. Solution Tree Press.

Heflebower, T., Hoegh, J., Warrick, P., & Flygare, J. (2018). *Teacher's guide to standards-based learning* (1st ed.). Marzano Resources.

Iamarino, D. L. (2014). The benefits of standards-based grading: A critical evaluation of modern grading practices. *Current Issues in Education*, *17*(2), 1-11.

Link, L., & Guskey, T. R. (2022). Is standards-based grading effective? *Theory into Practice, 61*(4), 406–417. https://doi.org/10.1080/00405841.2022.2107338

Marzano, R. J. (2006). *Classroom assessment and grading that work.* ASCD.

Marzano, R. J. (2010). *Formative assessment & standards-based grading*. Marzano Research.

O'Connor, K. (2018). *How to grade for learning: Linking grades to standards* (4th ed.). Corwin.

Peters, R., & Buckmiller, T. (2014). Our grades were broken: Overcoming barriers and challenges to implementing standards-based grading. *Journal of Educational Leadership in Action, 2*(2). https://doi.org/10.62608/2164-1102.1070

Reibel, A. R., Gobble, T., Onuscheck, M., & Twadell, E. (2025). *Pathways to proficiency: Implementing evidence-based grading* (2nd ed.). Solution Tree.

Schimmer, T. (2016). *Grading from the inside out: Bringing accuracy to student assessment through a standards-based mindset*. Solution Tree.

Townsley, M. (2019). Considering standards-based grading: Challenges for secondary school leaders. *Journal of School Administration Research and Development, 4(*1), 35–38. https://doi.org/10.32674/jsard.v4il.1941

Townsley, M. & Wear, N. (2020). *Making grades matter: Standards-based grading in a secondary PLC*. Solution Tree.

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*(3), 76–86, https://doi.org/10.1080/00098655.2024.2357615

SHAWN CARNEY, MEd, is a High School Science Teacher and serves as the AP Physics Course Captain in Fargo Public Schools. His major professional interests lie in the areas of standards-based grading, inquiry-based learning, and effective PLCs. Email: carneys@fargo.k12.nd.us

TERESA SHUME, PhD, is an Associate Professor and serves as the Director of Teacher Education in the School of Education at North Dakota State University. Her major research interests lie in the area of inclusive science teacher education. Email: teresa.shume@ndsu.edu

Manuscript submitted: ***September 7, 2025***
Manuscript revised: ***December 21, 2025***
Manuscript revised: ***July 2, 2026***
Accepted for publication: ***July 6, 2026***

***Special Issue**: Exploring Current & Future Grading Practices in K-12 Schools*
Research Article

JSARD
open access | est. 2016
Volume 11, Issue S1 (2026), pp. 131-141
Journal of School Administration Research and Development
ISSN: 2470-8496 Print/2470-850X Online
ojed.org/JSARD

De-weaponizing Grades Through Teacher Tinkering and Student Insight

Lance Piantaggini
University of Massachusetts Amherst, USA

ABSTRACT

When high-stakes assessments increase pressure and anxiety for students and teachers, they become weaponized. Jack Schneider and Ethan Hutt (2023) identified the permanence of grades as the root cause of weaponization in the American grading system. In this article, the author documents how they de-weaponized grades in their 9th-grade urban public school Latin classroom by using ungraded portfolios combined with self-grading. Seeking evidence of student experiences with grading throughout the school and looking to identify other practices that de-weaponize grades, the author conducted an exploratory qualitative study, blending a document analysis of 43 teachers' syllabi with in-depth interviews involving three 18-year-old senior class student participants. Participants shared negative experiences with procedures and policies related to grade permanence, including assessments lacking feedback and grading based on completion, accuracy, and effort. These procedures and policies appeared in 74% of teacher syllabi, showing that it was common practice to weaponize grades. Participants each suggested practices that could address grade permanence, including flexible deadlines, retakes, self-grading, and ungrading. These practices appeared in very few teacher syllabi, underscoring the need for de-weaponizing grades. This paper both presents solutions for de-weaponizing grades and shows how practitioners can engage in small-scale research at their schools to support reform efforts by using local, empirical findings.

***Keywords*:** grading, ungrading, self-grading, portfolio assessment

In their book *Off the Mark: How Grades, Ratings, and Rankings Undermine Learning (but Don't Have To)*, authors Jack Schneider and Ethan Hutt (2023) tracked the development of the American grading system to its current problems. They identified one problem as "weaponization," referring to high-stakes assessments that increase pressure and anxiety for students and teachers (p. 161). The root cause of weaponization is the permanence of standardized test scores and grades. Therefore, classroom practices that make grades "overwritable" serve as teachers' and students' main defense towards de-weaponizing grades (p. 232).

In Part I of this article, the author identifies weaponized grading practices found throughout the literature and various attempts to address grade permanence. In Part II, they document the overwriting of grades in their 9th-grade public school Latin classroom by using ungraded portfolios and self-grading. Part III features a small-scale exploratory study involving 43 teachers' syllabi and interviews with three 18-year-old senior class student participants guided by the research questions: What are students' perceptions of teachers' assessment and grading practices? How prevalent were practices used to overwrite grades at the school?

Findings show weaponization present throughout the school and how student perceptions of teachers' assessment and grading practices were mostly negative. Finally, the author concludes with promising solutions for de-weaponizing grades, based on students' suggestions in the study and supported by the literature.

PART I: LITERATURE REVIEW

Weaponization of Grades

Weaponization of grades in classrooms involves assessments that increase pressure and anxiety for students through permanence. An assessment is considered permanent when a grade is recorded in the school's learning management system (LMS) and has no opportunity to be overwritten, such as through revising and resubmitting (Schneider & Hutt, 2023). Grades are literally documented on a student's permanent record, whether a series of "50s" and "60s" on individual assignments or a "D" for the overall grading term. Unfortunately, the literature shows that students become overly concerned with avoiding low grades out of fear of this documented failure (Beck et al., 1991; Butler & Nisan, 1986; Chamberlin et al., 2018). An assessment could also be considered permanent if students perceive it to have lasting effects, such as constant low scores as a reflection of permanent traits, resulting in low self-esteem and disengagement from learning (Black et al., 2004; Harlen & Deakin-Crick, 2002; Koenka, 2022; Koenka et al., 2021; Pulfrey et al., 2011; Schinske & Tanner, 2014; Shim & Ryan, 2005). Thus, practices that overwrite grades, such as revisions and retakes, portfolios, and self-grading, and that create an environment in which their effects are not considered lasting, can address this permanence.

Addressing Permanence

Revision & Retakes

Overwriting assessment results can be accomplished by students revising their work and retaking assessments, which scholars advocate. As Jung (2024) said, "denying chances to revise or try again undermines the very purpose of education" (p. 179). Similarly, Townsley (2022) wrote, "educators should provide students with multiple opportunities spread out over time, rather than a single point in time, to demonstrate their learning" (p. 87). New assessment results can replace old ones in the LMS after further instruction, revision, and/or retakes. Feldman (2023) recommends that these revisions and retakes replace old ones without penalty.

Portfolios

Portfolios typically involve students collecting their own work and providing rationales for their chosen evidence, encouraging more ownership of their learning (Mahmood & Jacobo, 2019; Wormeli, 2018). For example, Bulgakov Cooke et al. (2022) reported the status of implementation goals at a middle school using digital portfolios over two years. The authors found that 72% of students felt that they had a say about what happened to them at school when using portfolios to curate their own work. This was up from 57% the year before. Since students are more likely to add their best work to portfolios, there is a low chance that they contain assessments of poor quality that would otherwise have lasting negative effects. Challenges to implementing portfolios, however, include concerns about a lack of standardization across teachers and inconsistent scoring of student work (Koretz, 1998; Schneider & Hutt, 2023). Systems to ensure uniformity can address some of these concerns (see, for example, Black & Wiliam, 2018).

Self-grading

Grading is a summative function used to report achievement at the end of learning (Black & Wiliam, 1998; Shepard et al., 2018). Self-grading promotes responsibility and accountability, and it can disrupt power relations over what becomes permanent (Nieminen, 2022). Still, some have advocated against self-grading, instead recommending ungraded self-assessment (Bourke, 2018). For example, Andrade (2019) identified self-assessment as a self-regulating skill that must occur before self-grading, and that this skill must be developed through an ungraded, formative process, not a graded, summative event. Andrade's position is that formative self-assessment is more productive than summative self-grading. Andrade also noted that students' self-assigned grades tend to be higher than grades given by their instructors (see Strong et al., 2004; Tejeiro et al., 2012), yet some studies have shown no significant differences between students' and teachers' grades (Darrow et al., 2002; Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000; Gorichanaz, 2024; Guberman, 2021; Kitchen et al., 2006; Sadler & Good, 2006; Sanchez et al., 2017). Beyond self-grading's validity, Boud (1989) made a case for self-grading on the grounds of "expediency," freeing up teacher time to provide more feedback rather than grading (p. 22).

PART II: TINKERING WITH OVERWRITING GRADES

Tyack and Cuban (1997) identified teachers as key actors in reform, tinkering with their practices towards positive change. Schneider and Hutt (2023) suggested recalibrating systems rather than reinventing them. These authors share the idea that small adjustments to practice can have a big impact on teaching and learning. In this section, the author documents recalibrations made to their classroom systems while tinkering towards de-weaponizing grades.

Ungraded Portfolios

While revision and retakes are overwritable options when assessments are graded, the alternative is to not grade individual assessments at all. Thus, the author used portfolios as collections of ungraded learning evidence showcasing student understanding and performance. Ungraded portfolios, then, effectively bypassed the need for revisions or retakes since grades were not being recorded outside of progress reports and quarterly report cards. Furthermore, students could choose to add early work to their portfolio and replace it later, or to wait until they had a final, more polished product to add. In terms of addressing permanence, it was not so much that grades were overwritable: the assessments themselves were overwritten.

Students added work to portfolios each week and were given time at the end of class to do so. The teacher encouraged students to add some assignments, while other assignments were of the students' choosing. It is important to note that the use of portfolios did not hinder the formative feedback process. If anything, the author observed that portfolios helped the process by preserving written feedback, available for students to access and review in class. Furthermore, since reviewing work was part of the self-grading procedure, portfolios supported the development of self-regulated learning skills.

Self-grading

While Schneider and Hutt (2023) noted problems with large-scale portfolio use, the author's experience was different at the classroom level. One practice used in combination with ungraded portfolios that may have mitigated negative effects was self-grading. During each term, students graded themselves using holistic rubrics that listed broad descriptions of the characteristics defining each level of the standard. These rubrics were presented in a Google Form, in which students selected one of six grades keyed to each level and provided a rationale with supporting evidence (e.g., "My grade should be ___ because ___ shows that I ___"). Figure 1 is one of two holistic rubrics used throughout the year.

Figure 1

Holistic Reading Rubric

A LOT of growth *Texts feel super easy to read, and they feel **at or below** my reading level.*	100 (A+)
Growing *Older texts seem easier to read, and/or current ones **feel like** my reading level.*	**95** **(A)**
Staying the same *Older texts seem just right, and/or current ones feel **a little above** my reading level.*	85 (B)
Fading *Older texts seem a little hard to read, and/or current ones feel **above** my reading level.*	75 (C)
Really fading *Older texts seem hard to read, and/or current ones feel **far above** my reading level.*	65 (D)
Falling behind *Texts feel very hard to read, and/or I feel like I've never seen these words before.*	55 (F)

Note. The revised design of numerical values with midpoint cutoffs (i.e., 55, 65, 75, 85, 95) was used after the author observed how a previous grading scale (i.e., 50, 60, 70, 80, 90) became a source of unnecessary frustration for students who were stressed by corresponding letter grades below the midpoint with minuses (i.e., D-, C-, B-, A-). Furthermore, seeking a perfect grade point average (GPA) drove some students to focus on their grades more than on learning. Since grades above "93" were equal to a GPA of "4.0" in

the district, a "95(A)" achieved the maximum GP. The "100(A+)" on this revised scale represented achievement beyond district and classroom expectations.

Students tended to under-grade themselves more often than to over-grade, perhaps a sign of low confidence, as documented in traditional point-accumulation classrooms (Black et al., 2004; Chamberlin et al., 2018; Townsley, 2022). In the few instances when a student's grade and/or rationale misaligned with the teacher's records and observations, the teacher and student met to discuss the discrepancies. Differences were resolved by the teacher showing the student how their learning evidence met or did not meet expectations, and by recording the student's new grade, which now aligned with the teacher's.

In sum, the author de-weaponized grades by using an ungraded portfolio system combined with self-grading. No permanent grades were reported until the end of the grading term. Students had autonomy to replace early learning evidence with recent learning evidence, and their self-assessment and self-grading of evidence could be adjusted through discussion. From the author's perspective, these recalibrated practices were fluid and lacked the permanence that weaponizes grades. Yet questions remained: Were teachers de-weaponizing grades in other ways throughout the school? What did other students think about their teachers' assessment and grading practices? To address these questions, attention was given to the most important interest holders: students themselves.

PART III: EXPLORATORY STUDY

Purpose & Research Questions

The purpose of this exploratory qualitative study was to explore students' experiences with grading across the school and to identify practices to de-weaponize grades. This study was guided by the research questions: What are students' perceptions of teachers' assessment and grading practices? and How prevalent were practices used to overwrite grades at the school?

Setting & Participants

This study took place at a small, urban, public high school in New England. Participants were a convenience sample of students at the author's school who were not taught by the author and were able to give consent independently. An interest survey was sent to all 18-year-old seniors. Three of these students agreed to participate.

METHODS

Data Collection

Given the purpose of exploring participants' lived experiences, interviews were the most appropriate choice. These were conducted late in the fall semester during students' study hall to avoid disrupting class time. An interview protocol was developed in consultation with a university professor of qualitative methods (see Appendix A). Interviews lasted between 15 and 30 minutes.

Data Analysis

Interviews were thematically coded by the author in Google Sheets, using initial coding followed by focused coding (Saldaña, 2013). This coding was done inductively to explore new ideas free from preconceived ones. Initial codes were refined during focused coding to identify recurring patterns in the interviews representing common experiences across participants. After identifying shared experiences with grading, 43 syllabi across all 21 teachers were collected and reviewed. The grading policy section of each syllabus was thematically coded using a procedure similar to that used for the interviews. For these documents, however, a deductive approach was used, based on student responses, to determine the prevalence of de-weaponizing grades at the school.

FINDINGS

Focused coding resulted in six categories, shown in Table 1. Based on these categories, participant experiences were categorized into two themes: *negative experiences* and *wanting more ownership.*

Table 1

Interview Categories and Their Reference Counts

Categories	References
Motivation & Interests	130
Procedures & Policies	111
Grades & Performance Expectations	86
Navigating Demands of High School	42
Meaning of Grades	18
Social Impact of Grades	17

Negative Experiences

Students felt their learning was restricted by assessments lacking feedback and graded on completion, accuracy, and effort. Their descriptions evoked a sense of compliance rather than curiosity about learning. All three students used remarkably similar phrasing regarding the lack of feedback on their work. Lina questioned, "I don't know why I'm getting a C or something." Sable remarked, "There was just one C . . . I was like 'what did I do wrong?'" Nigel commented, "I took a test and then ended up getting a bad grade . . . 'Oh what did I do wrong?'" While letter grades provide information about achievement, they offer no actionable steps for students. When asked about the lack of feedback, Lina wished her teachers "had said things more like 'here's what you can do to practice more to make sure you don't get those problems wrong in the future.'"

Assessments graded on completion, accuracy, and effort were also associated with a lack of meaningful purpose. For example, Nigel described "being fed information," and Lina found a lack of feedback unhelpful:

> A lot of times . . . your teacher will just mark your answer wrong and that's your grade. If you're just grading them on the number instead of how much they understand it, it doesn't really help them. I think that's the base of how most grading is, like the number of things you got right or wrong.

Syllabi showing assessments graded on completion, accuracy, or effort were quite common in the school. Such methods appeared on 74% of the syllabi, including the science course Sable was taking. Table 2 shows the overview of the teacher syllabi document analysis.

Wanting More Ownership

The second theme was participants wanting more ownership of their learning. Students suggested practices that would overwrite grades, including flexible deadlines, retake opportunities, self-grading, and courses with an ungraded approach.

Flexible Deadlines

Sable spoke about allowing students to slow down when concepts become challenging or speed up when thorough understanding is reached:

> Yeah it might take longer. Like, you will learn one topic at a time and then you move on to the next topic. That would definitely raise a lot of grades. That's what I feel like most teachers are missing out on. Like, learning for yourself not for the teacher.

Sable's description suggests that flexible deadlines could address low grades resulting from slow learning rates. She was in a learning environment with mostly fixed deadlines, as evidenced by 35% of syllabi listing late-work penalties and none explicitly indicating the possibility of extensions.

Table 2

Practices Found in Teachers' Syllabi

Practice	# Course Syllabi	% Course Syllabi	# Teachers	% Teachers
Grading Completion, Accuracy, Effort	32	74	16	76
Undefined Standards	16	37	10	48
Late Work Penalties	15	35	8	38
Retakes	7	16	6	29
Self-grading	4	9	3	14
Mention of Grading Scale	2	5	2	10
Ungrading	2	5	2	10

Retakes

Sable also thought students should have more opportunities to learn. This included more ways to show improvement and involve teachers guiding students through greater understanding via reteaching:

> I feel like they should go like second rounds. I feel that's the purpose of learning…you know, like, "Oh, you did this wrong, but I'll give you another chance to redeem yourself, but I will re-teach you before I give you the chance to redeem yourself or have a retake." If there was more chances teachers give to students to learn more . . . like for redemption.

When asked about her experience with retakes and teachers reteaching content, Sable shared, "It doesn't really happen often, especially in some courses that are considered more severe, like AP classes, for example. Those are pretty big classes, so yeah." Syllabi data showed seven instances of retake policies, sometimes conditional (e.g., optional retakes if the first score is below 80), making retake policies an uncommon grading practice in the school (See Table 2).

Self-grading

Nigel wanted students to have more responsibility in determining their course grade. He acknowledged that this would require teachers to relinquish some control and require trust. Nigel also thought it would hold students accountable as they develop into self-regulated learners:

> I would say let the students grade themselves, but you don't know if they're gonna be honest or not . . . I think letting students take responsibility for what they've learned . . . to see how honest they are, and to see if they actually learned what was taught to them . . . if they're mature enough to take responsibility and show the teacher they can really do it, and show they learned or maybe they didn't learn so they'll give themselves a bad grade, right?

Syllabi data shown in Table 2 indicate that self-grading was an uncommon grading practice found in just three syllabi, ranging from partially to fully determining the course grade.

Ungrading

Ungrading is the elimination or reduction of grades throughout the learning process (Kohn, 2020; Sorensen-Unruh, 2020). Lina described how learning from mistakes in a classroom without grades best supported her learning. Mistakes were not only allowed but even welcomed by her teacher as part of class procedures:

> We were going over a test we did, and I got most of it wrong, which was fine because we were going over how to fix that. That's one thing I like about class; she doesn't grade us on . . . how many questions we got wrong, she grades us on our understanding . . . we review it and we go over and we learn to solve the problems . . . I just find that's a lot better than just like "oh this was wrong."

Syllabi data shown in Table 2 indicate that ungrading was rare in the school, found in only two syllabi. Furthermore, Lina was the only student suggesting a practice she had experienced.

DISCUSSION

While negative effects of grading on student motivation have been documented in the past (Beck et al., 1991; Butler & Nisan, 1986; Chamberlin et al., 2018), findings from the current study provide deeper insight into how students experience those negative effects through their responses. All students felt their learning was restricted by assessments graded on completion, accuracy, and effort. Townsley (2022) lists these factors as contributing to a "point-accumulation classroom" favoring assessments as gradebook events, rather than formative assessments designed to improve learning (p. 14). Assessments graded on completion, accuracy, and effort were found in 74% of teacher syllabi in this study, making the point-accumulation classroom the school's standard. These findings are similar to those of Harland et al. (2015), who found that "in some situations it was possible to experience hundreds of graded assessments in an academic year" (p. 528), with students focused on doing work for grades and points rather than for their own learning. Therefore, in the current study, the lack of clarity about how students could improve their learning due to insufficient feedback is unsurprising. Students found themselves in a grade-heavy environment and suggested ways to address the permanence of grades.

Sable's desire for retakes was strong. Her use of words like "redeem" and "redemption" gives us a glimpse into what a teenager can be experiencing as a result of a teacher's procedures and policies. Retake policies can harness the power to neutralize the onslaught of grades and have support in the previously mentioned literature (Feldman, 2023; Jung, 2024; Townsley, 2022). Furthermore, Nigel's suggestion of self-grading is supported by the literature (Darrow et al., 2002; Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000; Gorichanaz, 2024; Guberman, 2021; Kitchen et al., 2006; Sadler & Good, 2006; Sanchez et al., 2017).

Lina's experience in a classroom that welcomed mistakes and had no grades was positive. Ungrading has been described as "an attempt to minimize the damaging effects of the final grade" (Kohn, 2020, p. xvi), and a way of "purposely eliminating or minimizing the use of points or letters to assess student work" (Sorensen-Unruh, 2020, p. 140). Given the point-accumulation classroom as the school's norm, ungrading would address this permanence. Researchers recommend replacing grades with feedback (Gorichanaz, 2024; Koenka et al., 2021; Newton et al., 2020; Rapchak et al., 2023), though not all educators see ungrading as a worthwhile alternative. Davis (2023) was unconvinced that grades are the problem, citing results from her study comparing three grading methods for first-year undergraduate writing students: traditional grading, portfolio grading, and ungrading. For the ungrading approach, Davis had students self-grade at the end of the term. Davis' main argument was that summative self-assessments used to arrive at grades did little to build the metacognitive skills leading to self-regulated learners. Still, a single instance of summative self-assessment at the end of the grading term may be beneficial if it allows for grading to be reduced throughout the rest of the learning process, as Lina experienced. Dyer (2024) was also concerned about the negative impact of implicit bias in both the instructor and the student when determining final grades. While implicit bias is likely present in any grading approach, Dyer claimed that the impact would be greater if self-grading were used for final course grades. The logic behind this claim is understandable. If implicit bias skews how achievement is reported, it is best suited to low-impact assignments. Still, implicit bias issues must be weighed against other negative impacts in the current point-accumulation classroom, such as students experiencing weaponized grade permanence more often. An ungraded approach might not be immune to bias, as Dyer writes, yet its benefits might outweigh this potential drawback. More research is needed on the equitable outcomes of ungrading. If done well, this approach could address grade permanence, de-weaponizing grades to the greatest extent possible by limiting them until absolutely necessary, such as at the end of the grading term.

Sable had negative experiences with fixed deadlines set by curricular pacing, and syllabi referenced accepting late-work penalties. Bloom (1968) advocated that students learn at their own pace, with teachers using formative assessment to provide feedback until they achieve mastery. This ongoing assessment process had no deadlines and was personalized according to individual student learning rates. Hills and Peacock (2022) report findings from a study involving 43 participants in an undergraduate biology course. Every assignment had a deadline and a one-week extension option. Of the 301 possible extensions, students chose this option for just 54 (18%). This shows students opted for extensions only as needed, and what was needed varied from student to student. For example, nearly all students (95%) chose the extension for the one high-

stakes essay in the course. For the other low-stakes assignments, 40% of students chose only one extension, and 35% chose none. These findings illustrate the range of students' needs and how flexible deadlines gave students greater ownership of their learning. While teachers can reasonably expect to teach students time-management skills to meet deadlines, late-work penalties harm slower learners, and fixed deadlines have been shown to decrease performance and engagement (Miller et al., 2019). In the spirit of providing an equitable education to all, flexible deadlines could be considered the garrison of late work penalties and would answer the call-to-arms against weaponized grades. Concerns about flexible deadlines exist, such as "cautioning about an avalanche of work to grade at the end of the semester to student exploitation of this policy" and a lack of "emphasis on personal responsibility" (Kruger, 2023, p. 234). However, as Kruger (2023) stated, there is little empirical evidence that such concerns manifest when deadline extensions are offered. Therefore, schoolwide flexible deadlines could result in overwriting what would otherwise be a permanent grade representing timeliness, not achievement.

LIMITATIONS & FUTURE RESEARCH

Syllabi alone could not have captured any teacher's complete practices. The degree to which teacher syllabi matched what was happening in class is unknown. Future research could validate syllabi data with observations, teacher interviews, assessment samples, or surveys. The small student sample size and short interview lengths were also limitations. Future research could involve a larger student sample with longer, multi-session interviews.

The revised 6-point scale in Figure 1 has not been researched, though it falls within optimal scale levels found to reduce variability and error (Guskey, 2024; Lozano et al., 2008). Its design, however, was intended to address motivational issues. Future research should investigate whether this scale keeps a focus on learning by eliminating the distraction of minuses (i.e., F-, D-, C-, B-, A-) and student concerns with achieving the highest possible score of "100."

Finally, future research should investigate how student-suggested practices are used across multiple teachers, and whether combinations of these practices are more effective than others in de-weaponizing grades, such as the combination of ungraded portfolios and self-grading used by the author. Furthermore, practitioners are encouraged to conduct small-scale studies in their own buildings to explore the de-weaponization of grades.

CONCLUSION

The author sought evidence of student experiences with grading and wanted to identify practices in the school that de-weaponize grades. Students identified negative experiences with grading, including a lack of feedback and assessments graded on completion, accuracy, and effort. Weaponized practices that emphasized points and grades were found in 74% of syllabi. Grades constantly surrounded students, and their interview responses showed concern about grade permanence.

Participants wanted more ownership of their learning and suggested practices to overwrite grades, including flexible deadlines, retakes, self-grading, and ungrading. The latter two practices were used in the author's own classroom. Furthermore, the combination of ungraded portfolios with self-grading asked students to be more responsible, as Nigel requested, building important self-regulated learner skills. However, the suggested practices to address grade permanence were uncommon elsewhere in the school.

While the generalizability of these findings is limited, this study shows that grade de-weaponization was rare throughout the school, though the practices that students suggested could be promising solutions for grade de-weaponization by overwriting them, just as Schneider and Hutt (2023) suggested. Therefore, practitioners are encouraged to explore the solutions presented in this paper, including flexible deadlines, retakes, and ungraded portfolios combined with self-grading.

REFERENCES

Andrade, H. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*(87). https://doi.org/10.3389/feduc.2019.00087

Beck, H. P., Rorrer-Woody, S., & Pierce, L. G. (1991). The relations of learning and grade orientations to academic performance. *Teaching of Psychology*, *18*(1), 35–37. https://doi.org/10.1207/s15328023top1801_10

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, *86*(1), 8–21. https://doi.org/10.1177/003172170408600105

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Bloom, B. S. (1968). Learning for mastery. Instruction and curriculum. Regional education laboratory for the Carolinas and Virginia, topical papers and reprints, number 1. *Evaluation Comment*, *1*(2), 1–12.

Boud, D. (1989). The role of self-assessment in student grading. *Assessment & Evaluation in Higher Education*, *14*(1), 20–30. https://doi.org/10.1080/0260293890140103

Bourke, R. (2018). Self-Assessment to incite learning in higher education: Developing ontological awareness. *Assessment & Evaluation in Higher Education, 43*(5), 827–839.

Bulgakov Cooke, D., Paeplow, C., Carter, J., & Belton, T. R. (2022). *Middle school competency-based grading initiative: 2018-19 to 2020-21*. Wake County Public School System. https://eric.ed.gov/?id=ED624580

Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, *78*(3), 210–216. https://doi.org/10.1037/0022-0663.78.3.210

Chamberlin, K., Yasué, M., & Chiang, I.-C. A. (2018). The impact of grades on student motivation. *Active Learning in Higher Education*, *24*(2), 109–124. https://doi.org/10.1177/1469787418819728

Darrow, A.A., Johnson, C. M., Miller, A. M., & Williamson, P. (2002). Can students accurately assess themselves? Predictive validity of student self-reports. *Update: Applications of Research in Music Education, 20*(2), 8–11. https://doi.org/10.1177/875512330202000203

Davis, H. T. (2023). Ungrading: Self-Assessment, effort, and motivation. *Composition Studies*, *51*(2), 143–147.

Dyer, J. (2024, January 8). Ungrading has an equity-related Achilles heel. *Grading For Growth.* https://gradingforgrowth.com/p/ungrading-has-an-equity-related-achilles

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*(4), 395–430. https://doi.org/10.2307/1170205

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287–322. https://doi.org/10.3102/00346543070003287

Feldman, J. (2023). *Grading For Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms* (2nd ed.). Corwin.

Gorichanaz, T. (2024). "It made me feel like it was okay to be wrong": Student experiences with ungrading. *Active Learning in Higher Education*, *25*(1), 67–80. https://doi.org/10.1177/14697874221093640

Guberman, D. (2021). Student perceptions of an online ungraded course. *Teaching & Learning Inquiry*, *9*(1), 86–98. https://doi.org/10.20343/teachlearninqu.9.1.8

Guskey, T. R. (2024, April 29). Addressing inconsistencies in grading practices. *Kappan Online*. https://kappanonline.org/addressing-inconsistencies-in-grading-practices/

Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2015). An assessment arms race and its fallout: High-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, *40*(4), 528–541. https://doi.org/10.1080/02602938.2014.931927

Harlen, W, & Deakin-Crick, R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. EPPI-Centre, University of London.

Hills, M., & Peacock, K. (2022). Replacing power with flexible structure: Implementing flexible deadlines to improve student learning experiences. *Teaching and Learning Inquiry*, *10*. https://doi.org/10.20343/teachlearninqu.10.26

Jung, L. A. (2024). *Assessing students, not standards: Begin with what matters most* (1st ed.). Corwin.

Kitchen, E., King, S. H., Robison, D. F., Sudweeks, R. R., Bradshaw, W. S., & Bell, J. D. (2006). Rethinking exams and letter grades: How much can teachers delegate to students? *CBE—Life Sciences Education*, *5*(3), 270–280. https://doi.org/10.1187/cbe.05-11-0123

Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2021). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback. *Educational Psychology*, *41*(7), 922–947. https://doi.org/10.1080/01443410.2019.1659939

Koenka, A. C. (2022). Grade expectations: The motivational consequences of performance feedback on a summative assessment. *The Journal of Experimental Education*, *90*(1), 88–111. https://doi.org/10.1080/00220973.2020.1777069

Kohn, A.F. (2020). Foreword. In S.D. Blum (Ed), *Ungrading: Why rating students undermines learning. (and what to do instead)* (p. xvi). West Virginia University Press.

Koretz, D. M. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education: Principles, Policy and Practice, 5*(3), 309–334.

Kruger, J. S. (2023). Rethinking penalties for late work: The case for flexibility, equity, and support. *Pedagogy in Health Promotion*, *9*(4), 234–236. https://doi.org/10.1177/23733799231198778

Lozano, L., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*, 73–79. https://doi.org/10.1027/1614-2241.4.2.73

Mahmood, D., & Jacobo, H. (2019). Grading for growth: Using sliding scale rubrics to motivate struggling learners. *Interdisciplinary Journal of Problem-Based Learning*, *13*(2). https://doi.org/10.7771/1541-5015.1844

Miller, L. A., Asarta, C. J., & Schmidt, J. R. (2019). Completion deadlines, adaptive learning assignments, and student performance. *Journal of Education for Business*, *94*(3), 185–194. https://doi.org/10.1080/08832323.2018.1507988

Newton, J. R., Williams, M. C., & Feeney, D. M. (2020). Implementing non-traditional assessment strategies in teacher preparation: Opportunities and challenges. *Journal of Culture and Values in Education, 3(*1), 39–51. https://doi.org/10.46303/jcve.03.01.3

Nieminen, J. H. (2022). Disrupting the power relations of grading in higher education through summative self-assessment. *Teaching in Higher Education*, *27*(7), 892–907. https://doi.org/10.1080/13562517.2020.1753687

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, *103*(3), 683–700. https://doi.org/10.1037/a0023911

Rapchak, M., Hands, A. S., & Hensley, M. K. (2023). Moving toward equity: Experiences with ungrading. *Journal of Education for Library and Information Science, 64*(1), 89–98. https://doi.org/10.3138/jelis-2021-0062

Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1

Saldaña, J. (2013). *The Coding manual for qualitative researchers* (2nd ed.). London: Sage.

Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, *109*(8), 1049–1066. https://doi.org/10.1037/edu0000190

Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE-Life Sciences Education*, *13*(2), 159–166. https://doi.org/10.1187/cbe.CBE-14-03-0054

Schneider, J., & Hutt, E. L. (2023). *Off the mark*. Harvard University Press. https://books.google.com/books?id=CoG3EAAAQBAJ

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, *37*(1), 21–34. https://doi.org/10.1111/emip.12189

Shim, S., & Ryan, A. (2005). Changes in self-efficacy, challenge avoidance, and intrinsic value in response to grades: The role of achievement goals. *The Journal of Experimental Education*, *73*(4), 333–349. https://doi.org/10.3200/JEXE.73.4.333-349

Sorensen-Unruh, C. (2020). A stem ungrading case study: a reflection on first-time implementation in organic chemistry II. In S.D. Blum (Ed), *Ungrading: Why rating students undermines learning (and what to do instead)* (p. 140). West Virginia University Press.

Strong, B., Davis, M., & Hawks, V. (2004). Self-grading in large general education classes: A case study. *College Teaching, 52*(2), 52–57.

Tejeiro, R. A., Gómez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., & Emberley, E. (2012). Summative self-assessment in higher education: Implications of its counting towards the final mark. *Electronic Journal of Research in Educational Psychology, 10*(2), 789–812.

Townsley, M. (2022). *Using grading to support student learning* (1st ed.). Routledge.

Tyack, D. B., & Cuban, L. (1997). *Tinkering toward utopia: A century of public school reform* (Revised edition). Harvard University Press.

Wormeli, R. (2018). *Fair isn't always equal, 2nd edition: Assessment & grading in the differentiated classroom* (2nd ed.). Stenhouse Publishers.

LANCE PIANTAGGINI, EdS, is a PhD Candidate in the College of Education at the University of Massachusetts Amherst. His major research interests lie in the areas of classroom assessment, feedback, and grading. Email: lpiantaggini@umass.edu

Manuscript submitted: ***May 5, 2025***
Manuscript revised: ***November 10, 2025***
Manuscript revised: ***April 30, 2026***
Accepted for publication: ***June 19, 2026***

APPENDIX

STUDENT INTERVIEW PROTOCOL

Relevant Demographics

- Pronouns

Focused Life History

- What brought you to this school?
- What's your first memory of getting a grade?
    - What was that like?
    - What did that feel like?
    - Tell me about a time when you got a good/bad grade (opposite of what was shared above: use *their* terminology)

Experience

- How would you describe your academics in terms of learning?
    - How do grades align with that?

Reflection of Meaning

- Tell me what grades mean to you. What makes you want/not want to get good grades?
- If you were to change something about grading, what would it be?

Closing

- Is there anything you want to talk about that we didn't get to?