# Grading Policies

*Ethan L. Hutt*
*K-12 Education*
*Standards & Accountability*

## Introduction

Ask students to talk about their experiences at school, and it likely will not be long before they mention their grades. For better or worse, grades are a defining feature of modern schooling, just as they have been for more than 150 years. Given the centrality of grades to the work of schools, it is unsurprising that grading policies have been a persistent source of concern, critique, and reform. When grades and report cards were first introduced in American schools in the mid-19th century, they were intended to offer an improvement on current practices of student evaluation that were often sporadic, informal, and ephemeral. As famed school reformer Horace Mann complained at the time, "If superior rank at recitation be the object, then, as soon as that superiority is obtained, the spring of desire and of effort for that occasion relaxes."[1] What was needed instead, Mann argued, were regular report cards. These report cards, which he likened to merchants' ledgers, would communicate to students and their parents how continued attention and investment in education would pay off over the longer term.

However, no sooner had these practices become widespread in American schools than educators and reformers began raising concerns about the role of grades in schools—concerns that will sound all too familiar to us today. Educators complained that grades had succeeded in focusing students on their studies but for the wrong reasons. The desire to acquire grades and tokens of achievement had become an end in itself, with intrinsic motivations being crowded out by the pursuit of external rewards.[2] Reformers likewise complained that teacher grades were often inconsistent, unreliable, and inflated, as they rarely produced a normal curve and only sometimes aligned with the results produced by standardized assessments.[3]

In the face of these persistent critiques about the inadequacy of grading practices and the potential ill effects of grades on schools, Americans have never been at a loss for alternative proposals. Whether this meant shifting to pass/fail grading, replacing letter grades with narrative grades, or eliminating grades altogether in favor of standardized assessments or nothing at all, Americans have been experimenting with alternative models for almost a century. The failure of these persistent efforts to change grading practices on a wide scale highlights the multiple functions that grades serve in our school system and their general place in our schools and society. These elements should serve as important context when considering existing practices and weighing the potential benefits of proposed reforms.

The history of grading and attempts at grading reform in the United States makes clear that grades have developed to serve three interrelated purposes in our system:

(1) grades serve to motivate students in their school work;

(2) grades communicate important messages both in the short term to parents and students and in the long term to future audiences such as colleges and employers; and

(3) grades synchronize the disparate parts of our decentralized school system by providing a common way of communicating about student achievement.

While grades serve multiple purposes, it is important to note that these purposes can often be in tension with each other. For instance, grades communicate important messages about how students are developing, and we hope that these messages will motivate students to work hard. But students who have already figured out they will earn an A in the course may stop putting in effort to improve and students who decide they are already too far behind to improve their grade may be demotivated by a grade's message. The synchronization function of grades can also, sometimes, be in tension with effort to improve grades' communication function. Grades help synchronize our system because everyone knows, in a broad sense, what an A or a C means and schools (and other institutions) honor these assessments even as we acknowledge variation in school quality. Narrative grades, by contrast, that provide paragraph long descriptions of students' strengths and weaknesses, provide far more detailed messages about what students did in a class and what they might be able to do in the future. But therein lies the rub. While grades have clear ordinal ranks, the same cannot be said of any of the countless adjectives that might be used to describe a student's work ethic or work product. This can make narrative grades informative but hard to use for administrative purposes.

In addition to these core functions, the cultural resonance of grades is important. That is, the question of grading policies concerns not only technical matters of scales, reliability, or precision but also matters dealing with wider expectations about what schools "should" do.

In almost all instances, policy interventions into grading are intended to address a perceived problem with one of these core functions of grading, and when these interventions backfire, it is almost always because they either undermine one of the other core functions of grading or undermine cultural expectations about grading. For instance, as I explain below, perennial concerns about "grade inflation" stem from the dual concerns that "easy A's" undermine student motivation (i.e., students will work less hard because standards are low) and undermine the communication value of grades, as student grade point averages (GPAs) will no longer accurately inform students or future audiences about students' level of academic achievement. Similarly, efforts during the pandemic to enact alternative grading models (e.g., universal pass/fail) were met with considerable pushback from parents and students who considered a school year without grades to be anathema to the purpose of schooling, namely, to reward effort and to allow students to distinguish themselves from their peers.[4]

## Key findings

With these considerations in mind, it is possible to think concretely about existing research on the role of grades in our school system and to think about the possibilities for current and future reform. The key findings of this body of research are:

**Key finding # 1:** *Grades are positively associated with and predictive of long-term school and life outcomes. For all the flaws and critiques of grades, research consistently shows that, for instance, students who do better in high school are more likely to go to and succeed in college. These findings mean that grades remain a useful source of actionable information for parents, practitioners, and policymakers.*

**Key finding # 2:** *Students calibrate their effort in school based on achievement standards and to match their desired or anticipated grades. Though many people might prefer students were intrinsically motivated by school, research indicates that grades provide meaningful incentives for students, which*

*suggests the importance of grades and grading tasks with high standards and meaningful academic outcomes.*

**Key finding # 3:** *Average student grades have steadily increased over time, which has raised concerns about grade inflation, but the long-term effects of this trend are unclear. The potential for grade inflation to reduce the utility of grades is real and deserves continued study and attention.*

**Key finding # 4:** *Teacher-issued grades can be influenced by racial and gender bias. Grades are ultimately a subjective evaluation, which requires that we are attentive to the sources of bias and that we reduce the effects of bias in the structuring of students' educational opportunities.*

**Key finding # 5:** *Grades remain a target of numerous reform efforts, but the evidence on the effectiveness of alternative approaches to fulfilling the core functions of grading and addressing concerns about equity and bias remains unclear. Many reform efforts fail because they treat grading as a technical exercise and do not tend to the cultural dimensions of grading or to the substantive issues of what grades should reflect. These issues are not solved by adjustments in grading scales or rubrics, they are only addressed through meaningful and substantive engagement with the question of the purpose of grading.*

# Evidence

**Key finding # 1:** *Evidence on the predictive value of grades*

Grades provide multiple forms of information to students, parents, and other potentially interested parties such as college admissions officers. A student receiving an A in a class is likely to understand that grade as indicating a job well done and as indicating a level of command of the important knowledge, skills, and/or habits of mind in the course. If this were the only message to be conveyed, then we likely would not need standardized grading systems across our school system, as any note of encouragement or congratulations would suffice. However, one of the enduring values of issuing grades is that they serve not only as a snapshot communication about performance in the present but also as a prediction of the likelihood of academic success in the *future*. Indeed, the predictive value of grades has been an area of great interest to a wide range of audiences—students, parents, teachers, guidance counselors, admissions officers, and policymakers—and has been a major area of research for decades.

This research can be broadly characterized as providing strong evidence that grades, especially students' high-school grade point average (HSGPA), are predictive of future academic success and attainment.

Most likely because of the financial and personal stakes involved, the greatest interest in the predictive power of grades has been around questions of college success and persistence.[5] In one particularly notable study, researchers used data from 150,000 students enrolling in flagship public universities nationwide and in four public state systems to conclude that HSGPAs were strong predictors of students' college performance. Specifically, a one standard deviation (s.d.) increase in HSGPA was associated with an 11 percentage point (p.p.) increase in six-year graduation rates at the least selective schools and a 4.5 p.p. increase in six-year graduation rates at the most selective schools.[6] Likewise, scholars using data on graduates from Chicago Public Schools from 2006 to 2009 concluded that the relationship between HSGPAs and six-year graduation rates is "strong and consistent" across a range of student and school achievement levels.[7]

Despite the strong evidence of the predictive power of grades, there is a recurring debate about whether other information, specifically standardized test scores, might better predict future

achievement. The research is clear not only that students' HSGPA is a better predictor of future college success than standardized test (i.e., ACT/SAT) scores but also that students' test scores and HSGPA offer distinct information about their academic abilities. Hence, the two pieces of information are best used in tandem.[8] There has been a great deal of interest in operationalizing this information to guide student and guidance counselor decision making, notably with the introduction of the concept of "college and career readiness" in the Common Core State Standards.[9] Here, research stresses both the value of thinking in these terms and the need to recognize that college and career readiness is not a singular construct but will vary depending on student demographics, career aspirations, *and* institution of higher education type and selectivity.[10]

The general statement that GPAs predict future academic success can be further refined when we consider the curricular experiences of students in high school. While overall academic performance, as reflected in a GPA, is useful, successfully completing certain courses (i.e., achieving a passing grade) also provides important predictive information about student success. Research has highlighted the importance of more rigorous course-taking patterns in general, but completing more higher-level math courses increases both students' likelihood of pursuing higher education and their future earnings potential.[11] In particular, there is evidence that completing specific courses such as Algebra I may increase enrollment and achievement in high-school math coursework[12] and the likelihood of college enrollment[13], that completing Algebra II may increase the likelihood of a student attending a two-year college,[14] and that completing an Advanced Placement (AP) course may increase the likelihood of college attendance.[15]

However, research suggests a cautious approach to pursuing policies of curricular narrowing in an effort to push more students into certain "gateway" courses. Efforts to increase access to algebra coursework in middle school have found only modest[16] and, in some cases, negative results for future attainment.[17] Likewise, an experiment involving the expansion of AP science courses to schools that had not previously offered them did not lead to an increase in four-year college enrollment.[18] The general point is that there is nothing magic about specific course titles or course grades—course content can be diluted, and grades can be inflated (see key finding #3 below). Hence, schools, districts, and states must take care in making the necessary investments in people, materials, and organizations to secure the benefits of these reforms.

**Key finding # 2:** *Evidence on the relationship between grading standards and student effort*

Anyone who has been a student, parent, or teacher knows that students can be positively or negatively motivated by the prospect of course grades. Scholars have formalized this basic intuition in economic theory, highlighting that students will likely calibrate their academic effort based on the value that they place on academic grades and their perceptions of how hard they will have to work to earn a given grade.[19] This view of "strategic" students who calibrate their effort based on the perceived difficulty and payoff of achieving a particular grade has important implications for education policy. Notably, students may end up withholding effort from courses in which standards are set either too low (e.g., working only hard enough to obtain the desired grade) or too high (e.g., declining to put in effort if a grade is out of reach). If this is true, then efforts to raise academic standards, as has been the aim of US education policy since the start of the Cold War, may influence student effort in school. In general, research supports the view that students calibrate their effort based on achievement standards, with students generally benefiting from increases in academic standards.

Several studies have tried to evaluate the effects of higher standards on student grades and achievement. The primary challenge of this kind of work is that it requires an independent measure of student achievement beyond the teacher's grade to determine the relative rigor of a teacher's grading

standards. Scholars have used standardized tests and end-of-course (EOC) assessments to assess grading standards, which requires the assumption that teachers and tests are largely aligned in terms of the learning outcomes and constructs that they are evaluating. Using several years of data on 3rd–5th graders in one Florida county, researchers concluded that students at all levels but especially those at higher achievement levels benefited from having teachers with higher grading standards.[20] Notably, the study determined that academic performance in both reading and math improved and that students who had teachers with higher grading standards were less likely to have disciplinary problems in school.

A study using nationally representative data on American high schoolers reached a very similar conclusion: students in schools with higher grading standards saw a boost in their standardized test scores.[21] While all students appeared to benefit from the standards, students at the top end of the achievement distribution experienced higher growth than those at the bottom of the distribution. While the higher grading standards were associated with higher test scores, those scores did not translate into higher graduation rates or college attendance rates, and this was especially true for minority students. Most recently, a group of researchers evaluated the effects of teacher grading standards in Algebra I on students' EOC scores and performance in subsequent math classes.[22] Consistent with the prior studies, these researchers found that students who experienced teachers with higher grading standards had greater test score growth and that this effect persisted in subsequent math performance. Importantly, the study found that students in all racial/ethnic groups benefited from having teachers with higher grading standards, as did students attending schools serving high and low socioeconomic populations.

A state grading policy change in North Carolina provides additional insight into how students respond to changes in grading policy.[23] North Carolina has a standardized grading scale for all the high schools in the state. In 2014, the state legislature changed the grading scale from a 7-point letter grade range to a 10-point range (i.e., moving from 93 points being the lowest A to 90). The upshot of the policy is that obtaining any given letter grade is three points "easier" and that the threshold for passing a course is a full 10-points "easier," with the lowest D shifting from 70 points to 60 under the new policy. If students applied themselves to their schoolwork regardless of grade incentives, we would expect a uniform, mechanical shift upward in GPAs across the grade distribution as the automatic result of the new policy. Instead, researchers found differential responses, with nearly all of the increases in GPA being achieved by high-achieving students; students further down the distribution did not see an increase in GPA. This finding suggests that lower-achieving students responded to the more lenient grading policy by decreasing their academic effort, a possibility seemingly confirmed by the notable increases in school absences and chronic absenteeism among these students.

Taken together, these studies suggest that students are sensitive to the expectations and demands set for them by academic expectations and grading standards. Because students rationally evaluate and adjust their effort based on the value of the perceived outcome, it is important to recognize that changes in policy must be accompanied by appropriate supports. Simply raising standards or adjusting grading scales is unlikely to achieve the desired outcome if students do not believe that they can (or want) the changed standard.

**Key finding #3:** *Evidence concerning grade inflation*

Few topics related to grading receive more attention in the popular press than the issue of grade inflation.[24] In general, grade inflation, like price inflation, refers to the tendency of GPAs to increase over time in an artificial way. It is important, conceptually, to recognize that there is nothing inherently wrong or problematic with rising student GPAs.[25] It could be the case, for instance, that the increase reflects schools' fuller commitment to helping all students reach their potential and/or, following an increase in academic standards and expectations, indicates an overall increase in the average educational

attainment of American students. In these cases, the rise in GPA corresponds to a rise in students' underlying achievement. The perennial worry that drives attention to grade inflation is that grades are increasing *without* a corresponding increase in student achievement (i.e., students' knowledge and skills). If this were the case, then grades would lose their communication value to students and parents regarding how well a student is doing and to admissions officers or employers regarding what a student has accomplished and is likely ready to do in the future. This would also imply that teachers were holding students to lower standards to achieve the same grade, which, as explored in the prior section, would likely have a detrimental effect on student learning. Even conceding that rising GPAs might lose accuracy in communicating about the overall level of attainment, they might still convey important information about students' achievement *relative* to others in their high-school class or national cohort. However, there are concerns that increases in GPA will compress student achievement into narrower and narrower bands because GPAs are expressed on a fixed scale.

There is broad agreement that student GPAs have increased over time without a corresponding increase in standardized test scores, implying some level of grade inflation in American schools. However, the long-term impact of this trend on the communication value of grades remains to be seen.

Studies that have examined trends in student GPAs over time agree that average GPAs are on the rise in American schools.[26] To this general statement about trends, we can add two important details: GPAs appear to be rising faster at schools serving students from higher socioeconomic backgrounds,[27] and students at very high-achieving schools may have lower GPAs—a so-called "frog pond" effect—relative to their test scores.[28] In both cases, these findings are important context for the quality and interpretation of the message communicated by GPAs.

Whether these grade inflation trends pose a threat to the messages conveyed by grades turns almost entirely on the question of how well one believes that standardized test measures proxy for student learning in a given subject (e.g., EOC) or more generally (e.g., SAT/ACT, National Assessment of Educational Progress (NAEP) scores). If GPAs are capturing cognitive and noncognitive skills, then there is nothing inherently contradictory about a rise in GPA without a corresponding rise in academic test score measures. Indeed, there are strong arguments that grades do and should reflect noncognitive skills and that a GPA—as an aggregation of numerous, independent semester-long assessments by teachers in a range of subjects—offers a better, more holistic assessment of student achievement than a single, hours-long standardized test.[29] That said, the lack of increase in the academic measures to match the rise in student GPAs would imply that students are improving solely or primarily in these noncognitive areas. Many scholars are skeptical that this could be the case, arguing that it is far more likely that the rise in GPAs reflects other changes, such as a shift in teacher expectations or organizational pressure to reduce the number of failing students.[30]

It is too early to know whether we have passed a tipping point where the increase in GPAs has undermined the communicative and predictive power of grades. For now, the uncertainty underscores the traditional wisdom and practice of triangulating student achievement through GPA and standardized test score information, as the best evidence continues to suggest that both have independent and additive value for understanding student achievement (see key finding #1 above).[31]

**Key finding #4:** *Evidence for bias in teacher grading*

As subjective assessments of student work, teacher-assigned grades have the potential to reflect implicit or explicit biases—racial, gender, socioeconomic—that teachers may have toward their students. Unfortunately, researchers have documented that teachers do, in fact, possess a range of biases toward their students, including beliefs about their interest in school and their ability to learn and expectations

about their likely future attainment.[32] Although research in American schools on the effects of this bias on student grades is limited,[33] research has consistently shown that teachers have more negative perceptions of minority students than of their White student counterparts.[34] Likewise, researchers have found that teachers' beliefs about their students are linked with their allocation of resources and attention and with the way teachers interact with and address students.[35] In addition to the way that teacher beliefs may shape teacher behavior toward students and, in turn, their educational outcomes, there is research that suggests that students may internalize perceived biases in ways that affect their school performance. Students' perceptions of their teachers' or their schools' expectations may subject them to "stereotype threats" that affect their self-perception, effort, or achievement.[36]

Bias against students of color also results in the well-documented disparities in school discipline rates, with students of color being both more likely to be referred for disciplinary infractions and more likely to receive harsher punishments for the same behavior as that of other students.[37] Exclusion from the classroom and/or school for disciplinary infractions is strongly associated with lower academic outcomes for students.[38]

Teacher biases may also structure students' opportunities within the school. For instance, researchers have long noted the considerable disparities in the number of low-income students and students of color identified for gifted-and-talented education programs. While these disparities could be a function of underlying inequalities in educational resources and access, a study of a district that switched from teacher recommendations to standardized test score results for the identification of gifted-and-talented students saw a significant increase in the identification of previously underrepresented groups.[39] This finding implies that teacher bias may result in an inability or refusal to recognize giftedness in low-income and minority students. Indeed, researchers have found that students of color are less likely to be recommended for gifted-and-talented programs even when they meet the selection criteria for such programs.[40] By contrast, research suggests that greater diversity in the teacher workforce leads to higher rates of identification of Black students for gifted-and-talented programs[41]—a finding that points to a much larger literature on the positive educational outcomes associated with students having a teacher of the same race.[42]

In the face of strong evidence of teacher bias, researchers have explored three primary ways of mitigating or eliminating these effects. First, a great deal of research has explored the effects of reducing the harm of bias by increasing teacher diversity. Researchers have consistently found that having a same-race teacher improves student outcomes at all levels of schooling across a range of outcomes, including attendance, achievement, graduation and disciplinary referrals.[43] The exact mechanisms driving these results remain a focus of scholarly inquiry, but the best evidence suggests that increasing teacher diversity increases the presence of same-race, same-gender role models, raises teacher expectations for students, and introduces a greater diversity of perspectives and cultural resources for students to draw on in school.[44]

Second, there has been a concerted effort over the last two decades to improve teachers' sensitivity to difference by emphasizing the importance of culturally relevant and responsive pedagogies, and this emphasis has been associated with improved student outcomes and increases in students' positive beliefs about themselves, their teachers, and their schools.[45] Similarly, there has been a concerted effort to create interventions that will either increase teacher sensitivity and empathy for all students[46] or directly address teachers' implicit biases.[47] However, there are questions about the enduring effectiveness of implicit bias interventions over the long term.[48]

Finally, rather than focusing on the dispositions or beliefs of teachers, some scholars argue that the best way to reduce bias in teacher grades is to frame the task of grading. Specifically, providing teachers with

a grading rubric with more clearly defined evaluation criteria successfully eliminated racial bias in grading in a randomized experiment.[49] These findings comport with the general tendency of underspecified evaluative criteria to be filled in with personal preferences and bias.[50] Nonetheless, to understand whether grading rubrics effectively reduce grading bias across the full range of academic skills, subjects, and levels of schooling, more research is needed.

**Key finding #5:** *Prospects for grading reform*

Discontent over the way that grades structure classes, motivate students, and communicate about student achievement has led to calls for alternative grading schemes for decades. The variety and number of reforms are too numerous to systematically review here, but there are two important points worth considering, as they apply to the evaluation of any prospective grading alternative or grading reform.

First, most importantly, many of the arguments in favor of particular reforms turn on normative questions that cannot be addressed by research. For instance, to take an issue that has studded the sections above, research can establish the degree to which teacher grades correspond to other measures of student achievement such as EOC or ACT scores. However, that research cannot answer what the "right" amount of correspondence should be between the two. That is a policy question, and it is one that turns on what we want and value in each piece of information. One could imagine many possible scenarios in which one or the other measure (i.e., grades or standardized test scores) is considered redundant and eliminated (e.g., many countries worldwide have high-stakes standardized tests as the primary or even sole arbiter of educational opportunity) or scenarios in which any overlap between the two was considered unnecessarily redundant. Historically, grades and standardized tests have worked in tandem to accommodate our desire for local control, decentralization, and a national system,[51] but compelling arguments can and have been made that this legacy should be set aside for more standardization and centralized control.[52]

Similarly, the recently popular push for "no zeroes" grading policies can be motivated by an empirically true statement: the traditional 100–0 scale gives more space to failing grades (i.e., below 60) than to passing grades. However, this observation does not answer the question of whether a missed assignment should be treated the same (either categorically or numerically) by a grading scale as work that was attempted but fell well short of expectations.[53] The same issue applies to debates about whether "process points" (i.e., was the work completed on time?) should be included in student grades.[54] There is certainly an argument to be made that grades should reflect only the quality of the work, but equally compelling arguments can be made that completing tasks on time is itself an important skill and one that is valued in the workforce. Research can provide an empirical description of the outcomes of one or the other policy (e.g., whether some students achieve higher or lower GPAs) or even whether certain policies are associated with particular long-term outcomes (e.g., increased likelihood of college graduation, better labor market outcomes, or increased likelihood to vote). How to weigh the merits or tradeoffs associated with these policies and outcomes is ultimately not a research question but a political question.

Second, we must evaluate prospective reforms not simply on the basis of arguments about their technical or normative superiority. Rather, we must evaluate them in terms of the demands that they place on our school systems when implemented. Much as we would like to deny or ignore the organizational constraints of our system, many promising reforms run aground on these realities. For instance, the concept "balanced assessment" was introduced in 2001 in an effort to make assessments of student skills more authentic, rigorous, and timely.[55] By integrating teachers' classroom assessments into larger district and state systems of assessment, balanced assessment promises to bring teacher

assessments of student work and external standardized tests into greater alignment.[56] Two decades later, however, the human, organizational, and political resources necessary to implement this reform with fidelity have proven nearly impossible to muster or sustain in practice.[57] On a less grand scale, efforts to replace letter grades with narrative grades have been around for decades, but there has been limited adoption of narrative grades because of the time required to produce them and because of the difficulty and uncertainty of interpreting their meaning.[58] While small, elite schools can often sustain divergent practices, the prospect of widespread adoption is remote.

## Conclusion

Grades are and will remain an important feature of our school system. Their persistence in the face of criticism, critique, and worry speaks to the multifaceted role that they play in our system and their relative success in fulfilling this role. However, this does not mean that we can or should turn our attention away from questions about grading. As the preceding sections illustrate, only through vigilant monitoring can we be assured that grades continue to communicate accurately and effectively with the full array of audiences who depend on them. Likewise, only through continued examination and deliberate reform can we ensure that our grading systems align with—or become better aligned with—our normative beliefs about how our school systems should operate and how grades should, among other things, motivate students, describe achievement, and pursue equity for all students.

# Endnotes and references

[1] Mann, H. 1846. *Ninth Annual Report*. Dutton and Wentworth.
[2] For example, see Sumner, R. G. 1935. What Price Marks? *Junior-Senior High School Clearing House* 9(6), 340–344.
[3] For example, see Weld, L. D. 1917. A Standard Interpretation of Numerical Grades. *The School Review* 25(6), 412–421.
[4] Whether and how schools should adjust their grading scales during the pandemic were a subject of considerable debate. For example, see Appiah, K. A. 2020, May 12. Despite the Pandemic, My College Is Allowing Grades. Isn't That Unfair? *The New York Times*. https://www.nytimes.com/2020/05/12/magazine/despite-the-pandemic-my-college-is-allowing-grades-isnt-that-unfair.html; Zeller, A. 2020, May 14. Oregon Parents Knock New Pass/Incomplete Grading Policy. *The Portland Tribune*. https://www.koin.com/news/education/oregon-parents-knock-new-pass-incomplete-grading-policy/.
[5] Adelman, C. 1999. *Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment*. U.S. Department of Education, Office of Educational Research and Improvement; Adelman, C. 2006. *The Toolbox Revisited: Paths to Degree Completion from High School through College*. U.S. Department of Education; Allensworth, E. M., and K. Clark. 2020. High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions about Consistency across High Schools. *Educational Researcher* 49(3), 198–211; Bowen, W. G., M. M. Chingos, and M. S. McPherson. 2009. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton University Press; Geiser, S., and M. V. Santelices. 2007. *Validity of High-School Grades in Predicting Student Success beyond the Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes*; Hurwitz, M., and M. Welch. 2018. Comment. In *Measuring Success: Testing, Grades, and the Future of College Admissions*. Edited by J. Buckley, L. Letukas, and B. Wildavsky. Johns Hopkins University Press; Long, M. C., D. Conger, and P. Iatarola. 2012. Effects of High School Course-Taking on Secondary and Postsecondary Success. *American Educational Research Journal* 49(2), 285–322; Noble, J., and R. Sawyer. 2004. Is High School GPA Better than Admission Test Scores for Predicting Academic Success in College? *College and University* 79(4), 17–22.
[6] Bowen et al. (2009).
[7] Allensworth and Clark (2020).
[8] Ibid.; Bowen et al. (2009); Kobrin, J. L., B. F. Patterson, E. J. Shaw, K. D. Mattern, and S. M. Barbuti. 2008. *Validity of the SAT® for Predicting First-Year College Grade Point Average* (Research Report No. 2008-5). College Entrance Examination Board.
[9] Porter, A. C., and M. S. Polikoff. 2011. Measuring Academic Readiness for College. *Educational Policy* 26(3), 394–417.
[10] Klasik, D., and T. L. Strayhorn. 2018. The Complexity of College Readiness: Differences by Race and College Selectivity. *Educational Researcher* 47(6), 334–351.
[11] Adelman (2006).
[12] McEachin, A., T. Domina, and A. Penner. 2020. Heterogeneous Effects of Early Algebra across California Middle Schools. *Journal of Policy Analysis and Management* 39(3): 772–800.
[13] Adelman (2006); Nomi, T., S. W. Raudenbush, and J. J. Smith. 2021. Effects of Double-Dose Algebra on College Persistence and Degree Attainment. *Proceedings of the National Academy of Sciences* 118(27): e2019030118.
[14] Kim, Jeongeun, Jiyun Kim, S. L. DesJardins, and Brian P. McCall. 2015. Completing Algebra II in High School: Does It Increase College Access and Success? *The Journal of Higher Education* 86(4), 628–662.
[15] Chajewski, M., K. D. Mattern, and E. J. Shaw. 2011. Examining the Role of Advanced Placement Exam Participation in Four-Year College Enrollment. *Educational Measurement: Issues and Practice* 30(4): 16–27.
[16] Dougherty, S. M., J. S. Goodman, D. V. Hill, E. G. Litke, and L. C. Page. 2017. Objective Course Placement and College Readiness: Evidence from Targeted Middle School Math Acceleration. *Economics of Education Review* 58: 141–161.
[17] Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor. 2015. The aftermath of Accelerating Algebra: Evidence from District Policy Initiatives. *Journal of Human Resources* 50(1), 159–188; Domina, T., A. McEachin, A. M. Penner, and E. K. Penner. 2015. Aiming High and Falling Short: California's 8th Grade Algebra-for-All Effort. *Educational Evaluation and Policy Analysis* 37(3): 275–295; Nomi, T. 2012. The Unintended Consequences of an Algebra-for-All Policy on High-Skill Students: Effects on Instructional Organization and Students' Academic Outcomes. *Educational Evaluation and Policy Analysis* 34(4): 489–505.
[18] Conger, D., M. C. Long, and R. McGhee, Jr. 2023. Advanced Placement and Initial College Enrollment: Evidence from an Experiment. *Education Finance and Policy* 18(1): 52–73.

[19] Becker, W. E., and S. Rosen. 1992. The Learning Effect of Assessment and Evaluation in High School. *Economics of Education Review* 11(2): 107–118; Betts, J. R. 1995. *Do Grading Standards Affect the Incentive to Learn?* Department of Economics, University of California, San Diego.
[20] Figlio, D. N., and M. E. Lucas. 2004. Do High Grading Standards Affect Student Performance? *Journal of Public Economics* 88(9–10): 1815–1834.
[21] Betts, J. R., and J. Grogger. 2003. The Impact of Grading Standards on Student Achievement, Educational Attainment, and Entry-Level Earnings. *Economics of Education Review* 22(4): 343–352.
[22] Gershenson, S., S. B. Holt, and A. Tyner. 2024. Making the Grade: The Effect of Teacher Grading Standards on Student Outcomes. *Contemporary Economic Policy* 42(2): 305–318.
[23] Bowden, A. B., V. Rodriguez, and Z. Weingarten. 2023. *The Unintended Consequences of Academic Leniency*. Annenberg Institute at Brown University.
[24] For example, see Donahue, T. 2023, October 23. If Everyone Gets an A, No One Gets an A. *The New York Times*. https://www.nytimes.com/2023/10/23/opinion/grade-inflation-high-school.html; Grose, J. 2023, October 18. Lenient Grading Won't Help Struggling Students. Addressing Chronic Absenteeism Will. *The New York Times*. https://www.nytimes.com/2023/10/18/opinion/chronic-absenteeism-lenient-grading.html; Hess, F. M. 2023, September 7. Grade Inflation Is Not a Victimless Crime. *Forbes*. https://www.forbes.com/sites/frederickhess/2023/09/05/grade-inflation-is-not-a-victimless-crime/.
[25] Pattison, E., E. Grodsky, and C. Muller. 2013. Is the Sky Falling? Grade Inflation and the Signaling Power of Grades. *Educational Researcher* 42(5): 259–265.
[26] Gershenson, S. 2018. *Grade Inflation in High Schools (2005–2016)*. Thomas B. Fordham Institute; Godfrey, K. 2011. *Investigating Grade Inflation and Non-Equivalence* (Research Report 2011-2). New York, NY: College Board; Perkins, R., B. Kleiner, S. Roey, and J. Brown. 2004. The High School Transcript Study: A Decade of Change in Curricula and Achievement, 1990–2000. *Education Statistics Quarterly* 6(1/2): 1–11.
[27] Gershenson (2018).
[28] Allensworth and Clark (2020); Koretz, D., and M. Langi. 2018. Predicting Freshman Grade-Point Average from Test Scores: Effects of Variation within and between High Schools. *Educational Measurement: Issues and Practice* 37(2): 9–19.
[29] Allensworth and Clark (2020); Pattison et al. (2013).
[30] Adelman, C. 2004. *Principal Indicators of Student Academic Histories in Postsecondary Education, 1972–2000*. Washington, D.C.: U.S. Department of Education; Gershenson (2018).
[31] Schneider, J., and E. L. Hutt. 2023. *Off the Mark: How Grades, Ratings, and Rankings Undermine Learning (But Don't Have To)*. Harvard University Press.
[32] Dee, T. S. 2005. A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review* 95(2): 158–165; Gershenson, S., S. B. Holt, and N. W. Papageorge. 2016. Who Believes in Me? The Effect of Student–Teacher Demographic Match on Teacher Expectations. *Economics of Education Review* 52: 209–224.
[33] Mechtenberg, L. 2009. Cheap Talk in the Classroom: How Biased Grading at School Explains Gender Differences in Achievements, Career Choices and Wages. *The Review of Economic Studies* 76(4): 1431–1459.
[34] Tenenbaum, H. R., and M. D. Ruck. 2007. Are Teachers' Expectations Different for Racial Minority than for European American Students? A Meta-Analysis. *Journal of Educational Psychology* 99(2): 253.
[35] Farkas, G. 2003. Racial Disparities and Discrimination in Education: What Do We Know, How Do We Know It, and What Do We Need to Know? *Teachers College Record* 105(6): 1119–1146; Ferguson, R. F. 2003. Teachers' Perceptions and Expectations and the Black–White Test Score Gap. *Urban Education* 38(4): 460–507; Tenenbaum and Ruck (2007); Warikoo, N., S. Sinclair, J. Fei, and D. Jacoby-Senghor. 2016. Examining Racial Bias in Education: A New Approach. *Educational Researcher* 45(9): 508–514.
[36] There is a very large literature on stereotype threats. For example, see Beilock, S. L., E. A. Gunderson, G. Ramirez, and S. C. Levine. 2010. Female Teachers' Math Anxiety Affects Girls' Math Achievement. *Proceedings of the National Academy of Sciences* 107(5): 1060–1063; Steele, C. M. 1997. A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist* 52(6): 613–629; Steele, C. M. 2011. *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*. W.W. Norton.
[37] Barrett, N., A. McEachin, J. N. Mills, and J. Valant. 2021. Disparities and Discrimination in Student Discipline by Race and Family Income. *Journal of Human Resources* 56(3): 711–748; Shi, Y., and M. Zhu. 2022. Equal Time for Equal Crime? Racial Bias in School Discipline. *Economics of Education Review* 88: 102256.
[38] Anderson, K. P., G. W. Ritter, and G. Zamarro. 2019. Understanding a Vicious Cycle: The Relationship between Student Discipline and Student Academic Outcomes. *Educational Researcher* 48(5): 251–262; Chu, E. M., and D. D. Ready. 2018. Exclusion and Urban Public High Schools: Short- and Long-Term Consequences of School Suspensions. *American Journal of Education* 124(4): 479–509.
[39] Card, D., and L. Giuliano. 2016. Universal Screening Increases the Representation of Low-Income and Minority Students in Gifted Education. *Proceedings of the National Academy of Sciences* 113(48): 13678–13683.

[40] Ford, D. Y., T. C. Grantham, and G. W. Whiting. 2008. Culturally and Linguistically Diverse Students in Gifted Education: Recruitment and Retention Issues. *Exceptional Children* 74(3): 289–306; Grissom, J. A., and C. Redding. 2016. Discretion and Disproportionality: Explaining the Underrepresentation of High-Achieving Students of Color in Gifted Programs. *AERA Open* 2(1). https://doi.org/10.1177/2332858415622175; McBee, M. T. 2006. A Descriptive Analysis of Referral Sources for Gifted Identification Screening by Race and Socioeconomic Status. *Journal of Secondary Gifted Education* 17: 103–111.
[41] Grissom and Redding (2016).
[42] Dee (2005); Egalite, A. J., B. Kisida, and M. A. Winters. 2015. Representation in the Classroom: The Effect of Own-Race Teachers on Student Achievement. *Economics of Education Review* 45: 44–52; Gershenson, S., C. M. Hart, J. Hyman, C. A. Lindsay, and N. W. Papageorge. 2022. The Long-Run Impacts of Same-Race Teachers. *American Economic Journal: Economic Policy* 14(4): 300–342; Lindsay, C. A., and C. M. D. Hart. 2017. Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina. *Educational Evaluation and Policy Analysis* 39(3): 485–510.
[43] Gershenson, S., M. J. Hansen, and C. A. Lindsay. 2021. *Teacher Diversity and Student Success: Why Racial Representation Matters in the Classroom*. Harvard Education Press; Gershenson et al. (2022); Lindsay and Hart (2017).
[44] Gershenson et al. (2021).
[45] Paris, D., and H. S. Alim. (Eds.). 2017. *Culturally Sustaining Pedagogies: Teaching and Learning for Justice in a Changing World*. Teachers College Press; Gay, G. 2010. *Culturally Responsive Teaching: Theory, Research, and Practice*. 2nd ed. Teachers College Press. Ladson-Billings, G. 1995. Toward a Theory of Culturally Relevant Teaching. *American Educational Research Journal* 32(3): 465–491; Ladson-Billings, G. 2014. Culturally Relevant Pedagogy 2.0: A.k.a. the Remix. *Harvard Educational Review* 84(1): 74–84; Warren, C. A. 2018. Empathy, Teacher Dispositions, and Preparation for Culturally Responsive Pedagogy. *Journal of Teacher Education* 69(2): 169–183.
[46] McAllister, G., and J. J. Irvine. 2002. The Role of Empathy in Teaching Culturally Diverse Students: A Qualitative Study of Teachers' Beliefs. *Journal of Teacher Education* 53(5): 433–443; Warren (2018).
[47] Whitford, D. K., and A. M. Emerson. 2019. Empathy Intervention to Reduce Implicit Bias in Pre-Service Teachers. *Psychological Reports* 122(2): 670–688.
[48] Forscher, P. S., C. K. Lai, J. R. Axt, C. R. Ebersole, M. Herman, P. G. Devine, and B. A. Nosek. 2019. A Meta-Analysis of Procedures to Change Implicit Measures. *Journal of Personality and Social Psychology* 117(3): 522–559.

[49] 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.
[50] Rivera, L. A. 2016. *Pedigree: How Elite Students Get Elite Jobs*. Princeton University Press; Uhlmann, E. L., and G. L. Cohen. 2005. Constructed Criteria: Redefining Merit to Justify Discrimination. *Psychological Science* 16: 474–480.
[51] Schneider and Hutt (2023).
[52] Polikoff, M. 2021. *Beyond Standards: The Fragmentation of Education Governance and the Promise of Curriculum Reform*. Harvard Education Press.
[53] For example, compare the arguments in Buck, D. 2022, June 16. A "No Zeros" Grading Policy Is the Worst of All Worlds. *Flypaper*. https://fordhaminstitute.org/national/commentary/no-zeroes-grading-policy-worst-all-worlds; Reeves, D. 2022, June 23. Revisiting "the Case against Zero": A Response to Daniel Buck. *Flypaper*. https://fordhaminstitute.org/national/commentary/revisiting-case-against-zero-response-daniel-buck.
[54] Feldman, J. 2018. *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms*. Corwin Press.
[55] National Research Council. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. National Research Council.
[56] Marion, S. F., J. W. Pellegrino, A. I. Berman. (Eds.). 2024. *Reimagining Balanced Assessment Systems*. National Academy of Education.
[57] Polikoff, M. S., and E. L. Hutt. 2024. The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities. In *Reimagining Balanced Assessment Systems*. Edited by S. F. Marion, J. W. Pellegrino, and A. I. Berman. National Academy of Education. 17–47.
[58] Schneider and Hutt (2023).

## Suggested citation

Ethan L. Hutt (2025). "Grading Policies," *in Live Handbook of Research on Education Finance and Policy*, in Douglas Harris (ed.), Association for Education Finance and Policy, viewed MM/DD/YYYY, <WEB LINK>.