EQUITABLE GRADING PROJECT

Can We Trust The Transcript?

Recognizing Student Potential Through More Accurate Grading

Foreword

Guadalupe Guerrero, CEO, Partnership for Los Angeles Schools; Former Superintendent of Portland Public Schools

Grades matter a lot. They shape a student's journey through school, dictating the courses they can take – perhaps, more importantly, the courses they can't take – the sports they can play, and what they are able to do when they graduate. For many of the nearly 2 million students who dropout of high school each year, grades provided the final push, symbolizing failure in the classroom.

But what if those grades aren't accurate? Or, vary greatly from teacher to teacher? What if the failing grade in one teacher's class that convinces a student to drop out of school is actually a passing grade in another teacher's class? As the research in this report demonstrates, this isn't a hypothetical scenario. Each year, millions of high school students are receiving grades that don't represent what they actually know and have learned. They are too high or too low, or they include factors – like showing up on time, or extra credit – that have little to do with the subject matter a student has mastered. Sometimes, the grades are a better reflection of the student's ability to negotiate with their teachers, or a parent's ability to attend a school event.

Yet, they are presented as the truth and set expectations about what that student can accomplish. Those grades get sent home to parents. They are shared with other educators and school leaders. They get rolled up with other grades – which may also be inaccurate – on transcripts and turned into a grade point average that is then shared as part of college applications.

While grading is important and, as shown in this paper, in desperate need of improvement, education leaders often disregard efforts to reform and align grading practices as they contemplate strategies like new curricula or pedagogical approaches – for turning around or otherwise improving their high schools. If district leaders do contemplate taking steps to reform grading practices, it's often after the "real reforms" take place.

This is unfortunate. I've witnessed, during my time leading the Portland Public Schools, the powerful and impactful ways that the equitable grading practices that are spotlighted in this report – like adopting a zero to 4 (or 50% to 100%) grading scale or excluding non-academic factors like student behavior, extra credit or parent participation from student grades – can not only drive consistency and accuracy, but can also lead to better student outcomes. This isn't easy work. We wanted grades in our schools to be more than just numerical scores. We wanted them to convey student growth and mastery, and to align with our values of equity and fairness. But those changes wouldn't just happen on their own. Few teachers receive training on grading in their prep programs, and the practices presented in this report overturn long-held conventional grading practices.

While the work has been challenging, it has also been incredibly important. It was about our values as a learning community, and ensuring our vision for learning aligned with those values. Ultimately, the work ensures that we are being honest with ourselves and our students about their learning. And nothing is more important than that.

Executive Summary

Grades play a central part in our K-12 system of education. The grades a student receives can determine their placement into honors or remedial courses and eligibility for after-school sports. In high school, the stakes of grades are higher: graduation, college admission, scholarships, financial aid, and even employment. As important, grades shape a student's self-perception and identity as a learner, which inform potentially life-altering decisions about their future, like whether they go to college or drop out of school.

Yet grading practices seldom appear in school or district improvement plans. They rarely are considered a lever to further district priorities like college and career readiness, engaging and relevant instruction, or student well-being.

As a result, while districts are quick to invest in new curricula and or professional learning based on contemporary research and technological advances, most school districts and educators continue to use grading practices that were developed during the Industrial Revolution, and are out of step with the rest of their practice.

More importantly, and as explained in this report, these **traditional grading practices yield inaccurate results: the grades students receive – that then inform big decisions about their future – don't accurately convey what students have learned.** In this study of approximately 33,000 grades in the 2021-22 and 2022-23 school years, nearly 60% of students' teacher-assigned grades did not match the standardized test scores designed to measure students' content knowledge of those courses. Two-thirds of those inaccurate grades were "inflated"–meaning a student's grades were higher than their test scores–and this occurred more frequently for Black, Hispanic, and students from families who qualify for free or reduced-price lunch ("FRPL students") than for Asian, white, and non-FRPL students. Additionally, one-third of those grades were "depressed"– the grade a student received was lower than their understanding of the course content as measured by those tests–and this

occurred relatively evenly across all student populations.

Even if we acknowledge that test scores are not the perfect measure of knowledge for some students–whether because of the limitations of the test design or because students may have test anxiety or simply had an off-day–that cannot explain away that **60% of the teacher-assigned grades were inaccurate.** Plus, while researchers and educators have raised concerns about grade inflation, particularly in the wake of the pandemic, this study finds that grade depression occurs frequently enough to raise an important alarm: because grades are used to open or close opportunities, significant numbers of students who are academically ready for advanced opportunities are at risk of being denied those opportunities solely because of the inaccuracy of the grades they receive from their teachers. **Those students' grades actually prevent us from recognizing their academic readiness.**

The good news is that teachers can improve their grade-test score accuracy simply by improving how they grade. This report will identify the traditional grading practices that, despite teachers' commitment to grade accuracy, yield inaccurate grades, and then will show data demonstrating that when teachers use a set of improved grading practices (categorized as "equitable grading" or "standards-based grading"), their students' grades more consistently match standardized test scores compared to grades assigned by a control group of teachers, and do so across nearly all subgroups of students, including those historically underserved and most vulnerable.

In addition, the grades students receive when teachers use more equitable grading practices are not only more accurate, but they are also higher: teachers using improved grading assign fewer D's and F's, particularly to those same historically underserved students, thereby reducing achievement disparities. This, alongside an increase in matching standardized test scores, suggests that the **improved grading practices are tied not to only reduced student failures, but also to increased learning and greater academic success.** Improving how teachers grade, while often considered an ancillary element to teaching and learning, is in fact a critically important strategy for districts and schools to address the persistent racial and economic gaps in student achievement.

The findings in this paper come at a particularly important time in K-12 education. In the wake of the pandemic, district leaders, educators and policy-makers continue to wrestle with addressing the widening academic gaps alongside increased student absenteeism[1] and decreasing public confidence in teachers,[2] while attempting to build trust in a system that often communicates, through grades, false information about student learning. In exploring solutions to these challenges, this paper makes the case that, based on findings from the study of districts and schools where teachers have learned and implemented improved grading practices, education leaders should look to these grading practices as key levers for driving systemic improvements.[3]

Contents

About the Equitable Grading Project 6

Acknowledgments 6

Introduction 7

The Study 11

Grades and (In-)Accuracy 12

Why Are Grades So Often Inaccurate? 15

The Impact Of Improved Grading Practices 18

- Impact: Grade-Test Score Consistency 20
- Impact: Grade Distribution 21

Implications and Recommendations 23

Conclusion 25

Endnotes 26

Appendix: Methods for Grade-Test Score Consistency and Grade Distribution 28

About the Equitable Grading Project

Established in 2021 by Joe Feldman, founder of Crescendo Education Group, the Equitable Grading Project is a non-profit organization dedicated to supporting the research, knowledge dissemination, implementation and expansion of improved, more equitable grading and assessment practices nationwide to improve success for all students, particularly for those historically underserved.

Acknowledgments

We are grateful to the educators and contributors who informed our perspective and provided insights.

Erica Price Burns, Senior Vice President, Whiteboard Advisors

Liz Cohen, Policy Director, FutureEd

Guadalupe Guerrero, CEO, Partnership for Los Angeles Schools; Former Superintendent of Portland Public Schools

David Hawkins, Chief Education and Policy Officer, National Association of College Admission Counseling

Christopher Love, Manager of Coaching, Crescendo Education Group

Dr. Sonal Patel, Assistant Superintendent of Educational Services, San Leandro Unified School District

Evo Popoff, Senior Vice President, Whiteboard Advisors

Brandon Protas, Assistant Vice President for Alliance Engagement, Complete College America

Dr. Jeffrey Tooker, Superintendent, Placer Union High School District

EQUITABLE GRADING PROJECT

info@equitablegradingproject.org

Introduction

Imagine a ninth grade student – we'll call her Olivia. Olivia has just checked her end-of-semester grades online, and sees that in her Algebra class she has a 79.3%, a C+. With another two-tenths of a percent–79.5%, which would be rounded up to an 80%–she would have a B-, the minimum required to enter Honors Geometry. She pores over her scores throughout the 18-week semester–a total of over 6000 points across categories as varied as homework, quizzes, class discussions, tests, extra credit, attendance, and projects–looking for anything she could use to convince her teacher to increase her grade. Her semester grade–which should reflect everything she has learned and months of hard work–may come down to her ability to negotiate with her teacher. Unfortunately, she knows that these rituals of checking end-of-semester grades and endlessly searching and negotiating for points will only intensify when she becomes a junior and is applying to colleges.

On that Spring end-of-course state algebra exam administered to every student taking algebra, Olivia scored at the "C" level. Olivia wonders: Even if she successfully negotiates with her algebra teacher to give her the points that qualify her for a B, does she actually understand Algebra to be ready for next year's math course? She also wonders about her English class, where the relationship between her grade and test score is flipped: based on her low grade she may need to repeat that course over the summer even though she performed quite well on her state's end-of-course assessment. **Are her grades telling the truth about her readiness for the tenth grade, or not?**

Her friend Jack is taught by the school's other algebra teacher who uses the same electronic textbook, worksheets, and tests, but has different grading policies. For example, Jack's teacher subtracts five points for every day work is handed in late, while Olivia's teacher subtracts ten points–often equating to a full letter grade–for each day late. Jack's teacher also offers extra credit points for creating a poster and pastry for "Pi Day" (March 14), and drops each student's lowest quiz score from the grade calculation. So, even when she and Jack received the same score on the end-of-course Algebra exam–suggesting that they have identical understanding of the Algebra standards–**they will receive different grades simply because their teachers grade differently.**

While Olivia is not a real student, her experiences are not hypothetical. They reflect the lived experiences of millions of students – in particular high school students – across the country. And as demonstrated by Olivia's story, grades matter a lot.

Grades can fundamentally alter a child's life trajectory, either opening doors or closing them. They determine whether students can enroll in honors academic tracks or must repeat courses, whether they are eligible for extra-curricular sports or need after-school tutoring, and whether they graduate and receive financial aid and scholarships. Particularly for students from families with fewer resources and privileges, the grades a student receives profoundly influence their chances to realize the American Dream of climbing the economic ladder and having financial stability, or being denied that opportunity and are more likely to replicate inter-generational financial hardship and struggle.

For this reason, we depend on grades to accurately describe a student's understanding of what was taught in a course. When a counselor, parent or college admissions officer looks at a student's grade in U.S. History, they likely believe that the grade accurately reflects the student's knowledge of U.S. History and what they learned in the course.

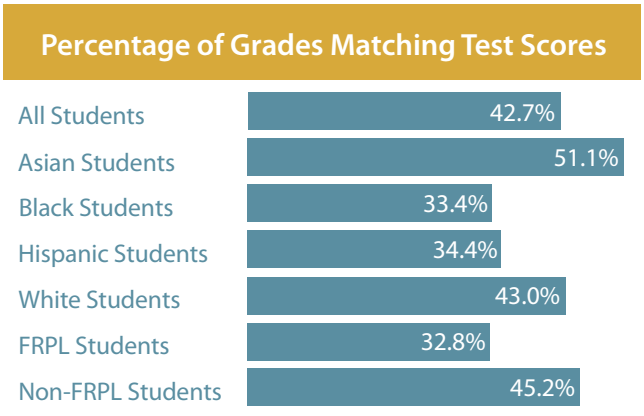
Indeed, a 2019 Learning Heroes survey[4] of nearly 3000 parents and guardians found that parents define their student's academic achievement through their grades. Parents rely on report card grades more than other resources, like classroom assessments or parent-teacher conferences, to determine if their child is achieving at grade level. Moreover, parents – in particular, parents of high school students – believe that their student's report card grades are the best source of information about their student's level of academic achievement.

Unfortunately, as demonstrated in this report, **a large percentage of student grades are not accurate reflections of what they know and have learned.**

Because grading happens entirely at the classroom level–by each teacher individually–there is no common checks and balances system to ensure that a grade is accurate. However, some states and districts administer a course-aligned assessment to provide an additional report of student learning. While both external tests and teacher-assigned grades are fallible, we would expect that teacher-assigned grades would largely match student scores on those external, course-aligned assessments. Unfortunately that isn't usually the case. Students' course grades often don't match their performance on standardized tests specifically designed to measure their understanding of that course content.

In a study of over 33,000 middle school and high school grades – discussed in more detail in this report – **nearly 60% of the grades did not match the expected assessment score, and rates of grade inaccuracy were highest among Black students, Hispanic students, and students from families with lower incomes ("FRPL").**

Even acknowledging the limitations of standardized tests, and accounting for students



Nearly 60% of the grades did not match the expected assessment score, and rates of grade inaccuracy were highest among Black students, Hispanic students, and students from families with lower incomes.

with test anxiety or who simply had a bad day when they took the test, this is an astonishingly high number.[5] Looked at another way, in this set of students, hundreds and perhaps even thousands of them received grades that did not match their course content knowledge as measured by a standardized test.

This disparity between students' grades and what they know can have a significant impact on their lives and may contribute to the low confidence in public schools and trust in teachers reflected in recent Gallup Polls.[6] In this study, the majority of the inaccurate grades were inflated–the grade was higher than the student's score on the assessment–suggesting that large numbers of students, and their families, were being told that they had achieved a level of academic understanding that they may not actually have had. In the future, those students may find themselves struggling in advanced courses or they may become one of the 40% of postsecondary students who are required to take remedial coursework.[7]

But this study also revealed a surprisingly high percentage of students with "depressed grades"–for whom their teacher-assigned grade is lower than their actual course knowledge as represented by a high test score, none of which can be dismissed because of test anxiety. The consequences of inaccurate grades are arguably more severe: these students, although qualified, are shut out from advanced opportunities, such as higher level math and science courses that are the pathways for competitive colleges or lucrative STEM careers.

The inaccuracy of grades is exacerbated by the variability of grading practices from district to district, school to school, and as we saw illustrated in Olivia's predicament, even from classroom to classroom. Perhaps nowhere is the accuracy of grades more important than in the college application process, but admissions officers identify grade inaccuracy as a significant concern in their determination of a student's college readiness. "Admissions officers understand that a lot of different factors, like assessment scores, extra credit assignments and classroom behavior may end up contributing to a student's grade, including factors that may have little or nothing to do with what a student knows or has learned," noted David Hawkins, the Chief Education and Policy Officer at the National Association of College Admission Counseling. **"Admissions officers would truly value more consistent high school grading practices that better represent the student's mastery of content when making admissions decisions, and those improved grading practices would also likely contribute to greater student success after they enter higher education,"** he continued.

Improving the accuracy of grades can not only address these challenges in K-12 and

college admissions, but may fundamentally improve the successful intersection of our K-12 to higher education system. Today, many institutions of higher education are reinstating mandatory standardized tests,[8] often citing declining academic rigor in student grades which can make it difficult to gauge if students are ready for college courses. "Higher education institutions rely on proxies for what students know – mostly standardized assessments – in making their decisions," explained Brandon Protas, Assistant Vice-President for Alliance Engagement at Complete College America. "Developing systems and practices that ensure that high school grades more accurately reflect a student's competence in their courses would not only reduce reliance on these proxies, which have their limitations, it would also open up possibilities for better vertical alignment between K-12 and post-secondary systems," Protas added.

Grades should be reliable. They should also be easy to understand. A parent or student or college admissions officer should be able to look at a grade and understand what that student has learned. But, as this study demonstrates, that often isn't the case. So why is it that grades are often inaccurate and difficult to understand, and therefore so untrustworthy?

This report explores both the challenges presented by traditional grading practices, and the opportunity to rethink high school education through the introduction of more equitable, standards-based grading. After a description of the study's findings on the incidence of grade-test inconsistency, this paper will explore how common grading practices themselves often lead to inaccurate grades, and why those practices exist and persist. This paper will then present findings from our research that includes teacher surveys and an analysis of teacher-assigned grades and student test scores to show the impact of research-based grading practices that improve the accuracy and fairness of grades.

Ultimately, this paper seeks to answer the fundamental questions: **"Can we trust our students' grades?" and, if we can't, "What can be done to restore that trust?"** In addressing these questions, the paper will also reflect on the powerful role that grading can and should play in efforts to improve K-12 systems of education–high schools in particular–and to reduce persistent achievement disparities.

The Study

This study of middle school and high school student grades is one of the largest studies ever on grading practices, and demonstrates the impact of specific, improved grading practices.

The 263 teachers[9] included in this study participated in one year of professional development to learn to implement improved, "equitable" grading practices. Each year of professional development included an extended Kickoff workshop, four 2½-hour workshops, and four 30-minute individual one-on-one coaching sessions led by Crescendo Education Group consultants, all former teachers. In the workshop series, teachers learned about the history of traditional grading practices and their negative impact on students, then learned about alternative, improved grading practices. As they began to implement the new grading practices, they received individualized coaching and engaged in cycles of action research in which they documented and shared their results with colleagues.

Four types of data were collected to investigate the impact of these grading practices:

1. Comparisons between students' Semester 2 grades and the test scores of those students in courses for which there was a matching externally-designed, standardized test. This data was collected in the Spring of 2022 before the professional development, and in Spring 2023 after the professional development (n = 33,383 test score-grade matchings / observations)
2. Semester 2 grades assigned by teachers in the professional development both before and after the workshop and coaching series (Spring 2022 and Spring 2023) (n = 58,751 grades)
3. Surveys of teachers (n = 1190 responses from 2019-2023)
4. Interviews conducted by Whiteboard Advisors on behalf of the Equitable Grading Project

The first two data sets were sent from districts directly to Elite Research Statistical Consulting in Austin, TX in compliance with FERPA and the districts' own privacy policies. In addition, that data included grades and test scores of students taught by teachers participating in the professional development and teachers in the same schools but who were not in the professional development, allowing for a quasi-experimental comparison. For additional information about Elite Research Statistical Consulting's methodology, see the Appendix. The third and fourth data sets were collected by Crescendo Education Group, which is affiliated with the Equitable Grading Project, in the course of their professional development. That data was analyzed by personnel who did not facilitate the professional development.

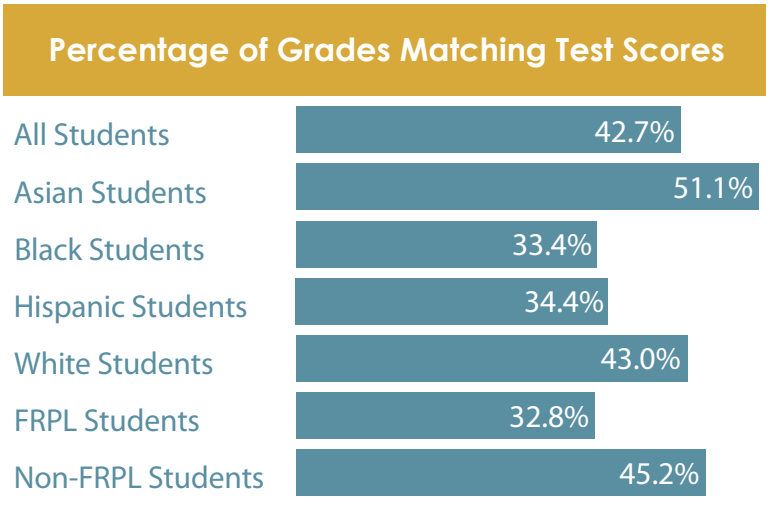
Grades and (In-)Accuracy

We would expect that a teacher's assigned grade for a course, particularly in second semester, would be a valid description of the student's cumulative understanding of that course content. Similarly, we would expect that a standardized test aligned to that course would yield a score that would also validly describe a student's understanding of that course content. And even with the limitations of any standardized test's reliability–that it is only a single assessment on a single day, and there are many reasons why a student may not perform well on that test, including test anxiety or just having a bad day–we would expect that most grades would match test scores for most students most of the time, what we call "grade-test score consistency."

However, **of approximately 33,000 students' grades across two years, only 42.7% matched the students' test scores.** Even if some of these pairings are dismissed because of the limitations of standardized tests, this suggests an astonishingly high rate of grade inaccuracy.

Even if we accept a rate of unavoidable mismatches between grades and test scores, we would still expect students of different ethnicities or economic groups to have similar rates of mismatches. However, **Black, Hispanic, and FRPL students had much higher rates of mismatches–meaning lower rates of grade-test score consistency–compared to their white, Asian, and non-FRPL peers.** For Asian students, whose grades were most frequently consistent with their test scores compared to other groups, still only about half of their grades were accurate.

Importantly, the 57% of grade-test score inconsistency did not occur equally in both "directions." For the purposes of this paper, "grade inflation" is defined as when the teacher-assigned grade is one or two achievement levels higher than the level of understanding indicated by the standardized test score, while "grade depression" occurs when the teacher-assigned grade is one or two achievement levels lower than the level of understanding indicated by the standardized test.



Of the 33,000 grades, over 40% were inflated, and grade inflation occurred more frequently for Black and Hispanic students than for Asian and white students, and was more frequent for FRPL students than for non-FRPL students. In the other direction, 16% of grades were depressed, and occurred relatively evenly across student populations.

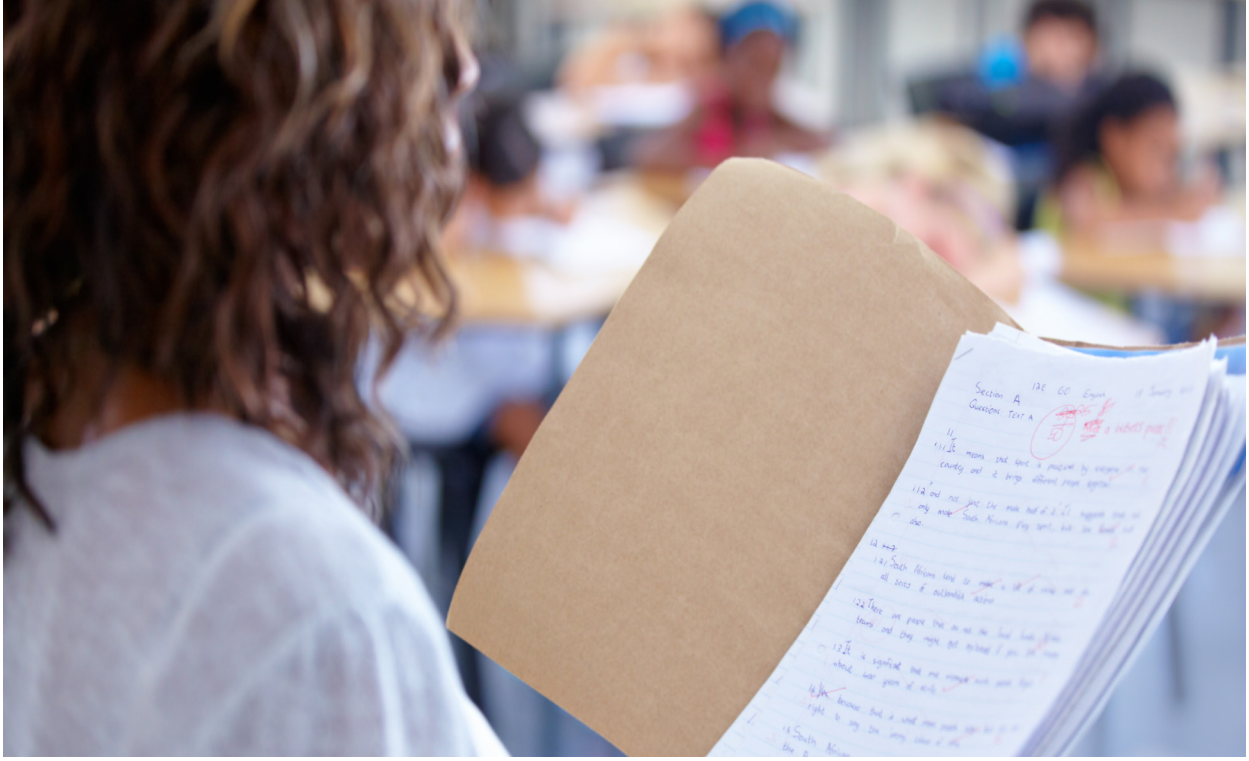
In real numbers, the results are sobering:

- Of the 33,383 grades assigned, over 13,000 grades were inflated, and almost **4800 of those grades were two letter grades higher than the student's test score.** Even if we account for students who are bad test-takers and the weaknesses of standardized tests, this still translates potentially into thousands of students from the study who received grades indicating they were more highly competent in the course content than they actually were. Those students may have been promoted or offered opportunities for which they were not academically prepared, and this occurred disproportionately for historically underserved student groups–those who may in fact have needed targeted support that their grades did not identify them as needing.

- Of the 33,383 grades assigned, approximately 5300 were depressed; in other words, nearly 1 out of every 6 grades assigned by teachers was lower than the student's test-indicated understanding of the course content. A significant number

Rates of Grade-Test Score Consistency, Grade Inflation, and Grade Depression: Spring 2022 and Spring 2023

*Percentages may not total 100% due to rounding



of students–**more than 1000–received grades in a course that were two letter grades lower than their assessment score** for that course content (for example, they received a teacher-assigned grade of a D even though they demonstrated B-level content understanding on the standardized exam for the course). Because grades unlock opportunities to students, this data suggests that hundreds, perhaps thousands, of students in this study may have been denied, or not even offered, opportunities that they were prepared and eligible for. At the very least, their grades told them they had lower competence in the course content than they actually had. We can imagine the psychological impact of this false message about their competence.

Grade inflation can occur for many complex reasons related to the interaction between a student and a test. A student's test score can be lower than a grade because of confusing or biased test questions, challenging testing conditions, or a student's test anxiety or a missed breakfast that can hamper their ability to show what they know. Grade depression, by contrast, has fewer explanations related to the test: students rarely score higher on a standardized test than their true understanding of the content. The primary causes are related to how the teacher chooses to grade.

Grade inflation can also be caused by the teacher's choices. Teachers may raise grades beyond students' true academic performance to alleviate pressure they feel from parents, administrators, or students. Teachers may assign higher grades than a student's true academic performance out of compassion and empathy for their students who, despite working hard, still underperform. Many other potential causes of grade-test inconsistency–both grade inflation and grade depression–are rooted in the teacher's curricular and instructional choices, pacing, and subject matter preparation.

However, this paper explores how common grading practices themselves–independent of a teacher's curriculum, instructional strategies, or students–cause both grade inflation and grade depression. When teachers implement specific improved grading practices described below, even when all other factors influencing inflation and depression remain constant, grades and test scores become more consistent. **Grades become more accurate when grading practices improve.**

Why Are Grades So Often Inaccurate?

Grades should be accurate and reliable. A parent or student or college admissions officer should be able to look at a grade and understand what that student has learned. But, as demonstrated above, that often isn't the case. **So why is it that grades are often inaccurate, and therefore so untrustworthy?**

First, traditional grading practices often capture more than what a student has learned. Teachers' grades commonly combine data from a diverse array of student performance inputs including tests, quizzes, homework, classwork, labs, projects, speaking during class discussions, presentations, and extra credit (which itself can range from doing extra work on a topic to bringing in classroom supplies). Moreover, as noted by Dr. Jeffrey Tooker, the Superintendent of Placer Unified School District, "traditional grades, far from simply conveying what a student knows, have been used in classrooms as a tool and incentive to modify student behaviors – like speaking without raising your hand or showing up late to class – or simply to ensure compliance or completion of assignments." While these may be important skills in some contexts, they generally are not course outcomes, and, as Tooker notes, "we have found in our district that there are more effective ways to address those issues than through a student's grades."

In addition, **traditional grades can encompass criteria that are not only unrelated to the content of the course, but unrelated to the student as well.** As Tooker reflected,"it wasn't unheard of, before we instituted new grading policies, to see students earn points toward their grade based on the engagement of their parents or caregivers." "Not only does a parent's attendance at Back to School Night have nothing to do with what a student knows," Tooker continued, "but you are effectively rewarding students with better grades based on their available resources – such as parents who have the professional flexibility to attend a school event."

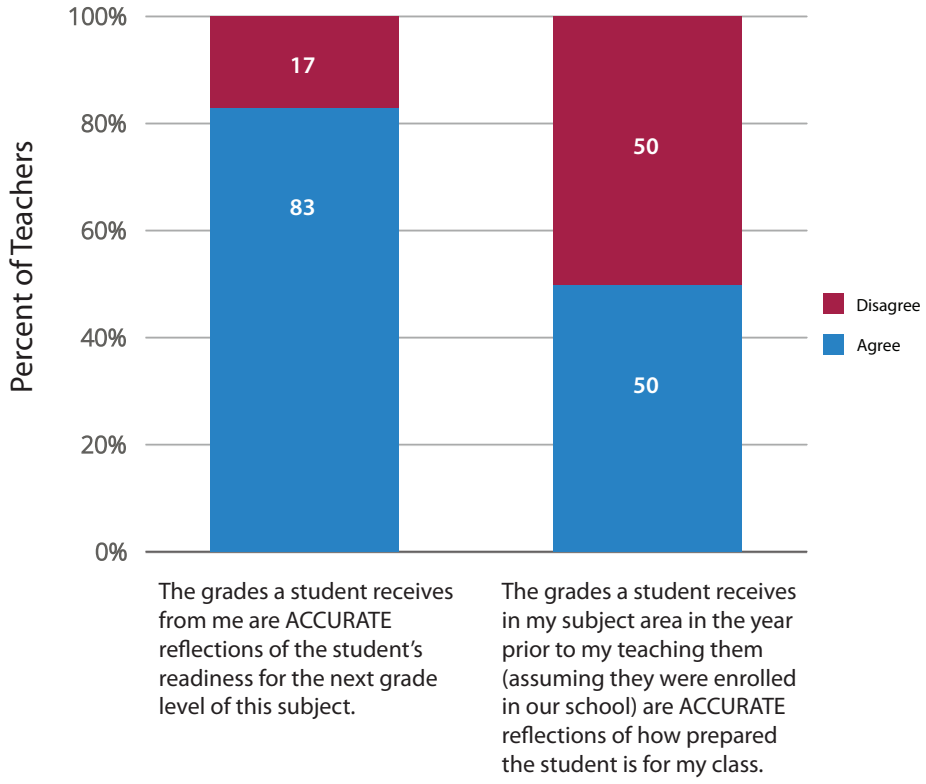
A teacher inputs this cornucopia of performances into a grading software's highly sophisticated algorithm, yielding a single percentage or letter grade. **These omnibus grades, in their attempt to capture all aspects of a student, end up conveying little meaning about the student's knowledge of the course outcomes.** Take the following example:

Grading Category and Weight	Letter Grade
Participation + Attendance (30%)	A-
Homework Completion (30%)	B+
Extra Credit (10%)	A+
Assessments of Content Understanding (30%)	D
Total Grade	B

This student has weak understanding of course content that clearly warrants remediation and support, but their high performance in non-academic areas results in an inflated grade that misrepresents their performance and conceals their academic needs. The reverse can also happen: a student with high academic understanding but low non-academic performance can receive a depressed grade that inaccurately communicates that the student is in need of remediation or unprepared, when in fact the students may be ready for additional academic challenges.

The problem isn't that teachers don't have the desire, or intent, to be accurate. In a survey we administered to nearly 1200 teachers, nearly every teacher (99%) reported that accurate grades are important, with three-quarters reporting that accurate grades are "extremely important." Ironically, teachers believe that their own grades were accurate – 83% agreed or somewhat agreed that their grades accurately reflect student learning and academic readiness – but that belief does not always extend to their colleagues' grades: nearly 50% of those same teachers doubted the accuracy of grades assigned by other teachers–teachers not in another district or school, but in their own school and department.

Teachers' Confidence in Accuracy of Their Grades and Their Colleagues' Grades



To be clear: teachers are not to blame for this grading variability and inaccuracy. Most teachers do not receive training in how to grade accurately. As noted by Dr. Sonal Patel, Assistant Superintendent of Educational Services at San Leandro Unified School District, "there is this assumption that teachers inherently know how to grade, so most don't receive professional development about grading best practices and we end up with highly variable grades across classrooms, schools and districts." Dr. Patel's observation is confirmed by research. In a survey of over two dozen teacher education programs (both university and alternative certification programs), while students studied such topics as the history of curriculum, content-specific pedagogy, and classroom management strategies, assessment literacy and the understanding of how to describe student performance was rarely if ever included in the curriculum.[10]

In the same survey of approximately 1200 teachers, **over half had no understanding of the history of grading, grading research, or strategies to grade that would improve accuracy or reduce subjectivity and biases. Nearly a third did not understand the mathematical structure of the 0-100% scale,** perhaps the most ubiquitous framework in common grading. According to Patel, "when we started the work of changing grading practices in our district, we found that over 95% of the teachers had never received any formal training in how to grade in their credentialing program, yet they also felt very passionate about their grading practices and their right to not have them altered."

Moreover, teachers rarely collaborate in grading with their colleagues–who likely also haven't received training on grading practices in their teacher preparation. Left to their own devices, with limited access to contemporary grading research and strategies, teachers have little choice but to fall back on their own experiences with grading, replicating how they were graded in high school and adopting traditional grading practices that have been in place for generations. After all, as Tooker noted, "everything in education has changed, except the way we grade."

Given this lack of education on effective and accurate grading practices, it shouldn't be a surprise that grading practices can vary greatly from classroom to classroom. As Patel recounted, "before we changed our practices, if we had three different teachers teaching 9th grade biology, they might have three very different approaches to grading resulting in three different grades for students with the same level of mastery."

If grading is so variable that a student's grade can depend on their teacher's individual approach, and teachers don't have confidence in the accuracy of the grades assigned even by their own colleagues, then it may be that **though grades are relied upon for decisions made by the student, their family, their school, and potential admissions officers and employers, those grades shouldn't be trusted.**

> "Everything in education has changed, except the way we grade.
>
> Dr. Jeffrey Tooker
> Superintendent, Placer Union High School District

The Impact Of Improved Grading Practices

Can we restore trust in teachers' grades? Can grades be more valid, accurately describing a student's academic understanding of their courses? Fortunately, specific grading practices, supported by contemporary research and the evaluated impact of those implemented practices, suggests that the answer is "yes."

Improved grading practices, sometimes called "standards-based grading" or "mastery grading," have been identified among researchers such as Susan Brookhart, Lee-Ann Jung, Robert Marzano, Ken O'Connor, Douglas Reeves, and Rick Wormeli. Most recently, Joe Feldman, author of several publications including Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms, has incorporated culturally-sustaining pedagogy research to identify "equitable grading" practices that were taught to the teachers in this study and surveys.[11]

Equitable Grading Practices

- Using a 0-4 or 50-100% grading scale
- Only report a student's academic understanding of course content at the end of their learning, and therefore excludes formative assessments (including classwork and homework) in the final grade
- Exclusion of a student's non-academic or behavior performance, including extra credit and participation, from the grade
- Non-grade-based consequences for late work or cheating/copying
- Rubrics and proficiency scales to evaluate assessments
- Only an individual student's performance, not a group's performance
- Offering students multiple opportunities to demonstrate their understanding, including redos and retakes, and utilizing a range of assessments

Equitable grading practices are intended to impact the grade in several ways:

- They ensure the grade reflects only a student's knowledge of the course standards and outcomes
- They apply a numerically linear scale that results in calculation outcomes that are more mathematically sound and accurate
- They make the grade less dependent on circumstances outside a student's control, including their environments and resources
- They reduce the impact of a teacher's subjective and potentially-biased judgments of a student's behaviors

Educators and leaders have noted that these practices can have a profound impact in their classrooms, creating a greater focus on academic knowledge and increasing academic rigor. As Tooker noted, "The practices have changed everything we do, because now when our teachers are working with our students they are talking about learning, not about points." As a result, **"We saw a closing of the achievement gap, especially with students of color who received fewer D's and F's, but also saw fewer A's because these practices reduced grade inflation," continued Tooker, "and we are seeing that those high achieving students are better prepared when they get to college, because they know what they know and what they still need to learn."** As Patel notes, "we have seen a culture shift with our teachers – and more consistency in how they graded students – as they realized that they had been grading busy work or over-inflating grades because they were incorporating extra-credit or factors that had nothing to do with the coursework."

Impact: Grade-Test Score Consistency

How do improved, equitable grading practices affect a grade's accuracy? Students taught by teachers trained in the use of equitable grading increased their rate of grade-test score consistency–meaning a statistically significant higher likelihood of a match between a student's teacher-assigned grade and the student's score on that course-aligned standardized test–by 8.1%, double the improvement of grade-test score consistency of students taught by teachers in the control group who did not receive the professional development. This increase in grade-test consistency among students taught by teachers using equitable grading exceeded that of the control teachers' students among FRPL and non-FRPL students as well as all racial subgroups except Black students. While the anomalous result for Black students is certainly worthy of further study,[12] this data suggests that the students who have the greatest inconsistency between grades and test scores experience the largest corrections when their teachers use equitable grading practices.

In addition, teachers using equitable grading practices decreased the rate at which grades were inflated by 7%, and did so for every subgroup of students except Black students.

Interestingly, both the control teachers and the equitable grading teachers had an increase in grade depression from 2021-22 to 2022-23, which possibly represents an "over-correction" after 2019-20 and 2020-21 pandemic-era adjustments to grading. However, while the control group's rate of grade depression increased by 21.4%, the rate of grade depression among teachers trained in equitable grading practices increased by only 4.7%.

While more research is necessary, the data indicates that **improved, more equitable grading practices increase grade-test score consistency, and thereby make teacher-assigned grades more accurate and valid.**

Percentage Improvement in Grade-Test Score Consistency from Semester 2 2021-22 to Semester 2 2022-23

	Among Students Whose Teachers Received Training in Equitable Grading Practices	Among Students of Control Teachers Who Did Not Receive Equitable Grading Training
All Students	+8.1%	+3.9%
Asian Students	+13.3%	+2.5%
Black Students	-25.9%	+2.4%
Hispanic Students	+11.2%	-4.4%
White Students	+5.0%	-0.5%
FRPL Students	+9.3%	-1.2%
Non-FRPL Students	+17.5%	+5.3%

Impact: Grade Distribution

If teachers assign more accurate grades after professional development in improved, more equitable grading practices, how does that training affect the rate at which teachers assign different grades? Using an expanded data set of 58,751 grades, external evaluators compared the rate of A-F grades assigned ("grade distribution") by teachers before the professional development (2021-22) and after the professional development (2022-23).

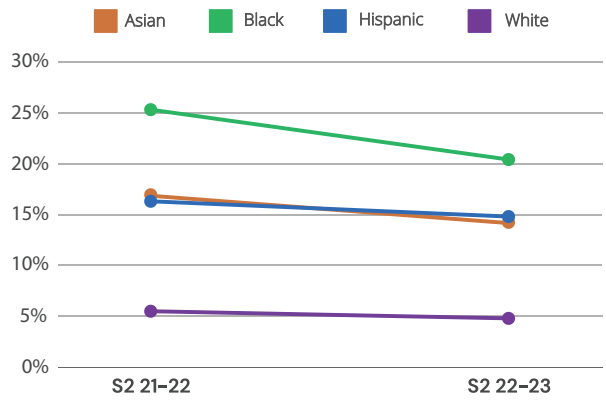
After receiving professional development to learn and implement improved grading practices, teachers decreased the rate at which they assigned D's and F's as well as A's.

Grade Distribution Before and After Professional Development (Semester 2 2021-22 & Semester 2 2022-23)			
	S2 21-22	S2 22-23	Difference
A	48.5%	47.9%	-0.6%
B	24.1%	26.9%	+2.8%
C	15.0%	15.5%	+0.5%
D	7.4%	6.1%	-1.3%
F	5.0%	3.5%	-1.5%

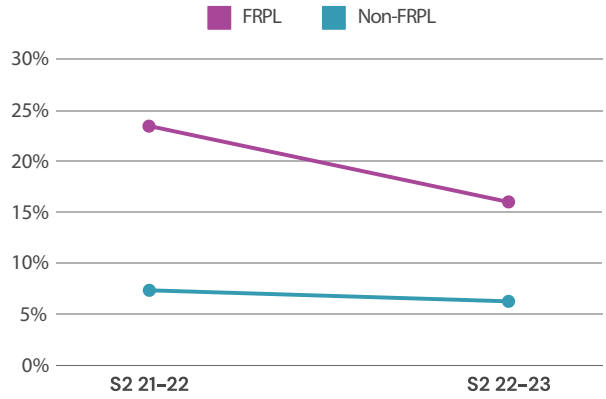
It's important to remember that many of these same teachers were assigning course grades that were significantly more consistent with their students' standardized test scores for that course content (+8%). **This suggests that students' grades were becoming more accurate, or students were learning more, or both.**

The changes in rates of assigned grades reduced achievement disparities by both race and family income. The graphs below show the changes in the rate of A's and D's / F's assigned by teachers pre- and post-training in equitable grading practices, and the narrowing of achievement gaps as represented by student grades.

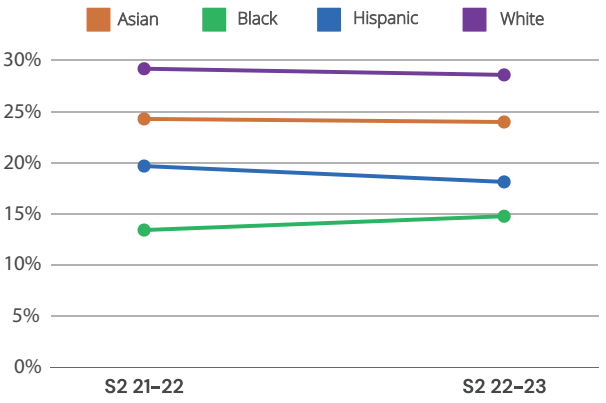
Rate of D/F Grades Assigned: Ethnicities



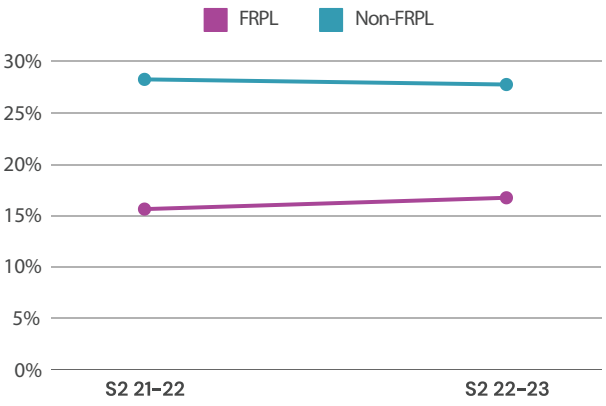
Rate of D/F Grades Assigned: Income



Rate of A Grades Assigned: Ethnicities



Rate of A Grades Assigned: Income



Importantly, this grade distribution data was also analyzed for marginal probabilities, with similar results, bolstering the likelihood that the changes in grades were tied to teachers' training and implementation of equitable grading practices.

This study also found a statistically significant relationship between a teacher's use of the practices and grade-test score consistency and fewer D and F grades. In other words, the greater a teacher's implementation of the improved practices (measured by self-reported frequency of using the practices), the greater likelihood of grade-test score consistency and of a reduction in assigned D and F grades. **This suggests that as teachers become more familiar and experienced with using the practices, and use more of them, there is an increase in grade accuracy and higher achievement.**

Finally, we found that when teachers learned about and used improved, more equitable grading practices, not only did it lead to their grades becoming more accurate and raising passing rates; this investment in their professional skills impacted their decision on whether to remain a teacher in their district or school.

Teacher retention is a constant concern for administrators, a concern heightened since the pandemic. A National Center on Education Statistics report showed that in the 2021-22 year, 8% of teachers left the profession and another 8% switched their school, and the National Education Association reported that over half of its teacher members were considering leaving the profession.[13] In our survey of nearly 1200 teachers, including the 263 teachers in the analysis of student grades, who experienced the professional development to learn to implement equitable grading practices, nearly 40% indicated that they were more likely to remain in their school or district specifically because of that training.[14] While there are many factors influencing teachers' career choices, this data suggests that **supporting teachers to improve their grading practices–addressing this crucial gap in their professional training–may increase teachers' job satisfaction and provide a strategy that both benefits students and reduces teacher attrition.**

Implications and Recommendations

These results indicate that when teachers are taught to use improved, more equitable grading practices, their grades become more accurate, reduce achievement disparities, and promote higher student achievement.

Implications

Implication #1: Improved grading is crucial to the integrity of our classrooms and schools because accurate, equitable grading both opens opportunities for prepared students and targets remediation for those who actually need it. Consider the findings of this study: for teachers receiving professional development in improved, equitable, grading practices, their grades more accurately described students' understanding of the course content. At the same time, the rate at which students received F's decreased nearly 30% (from 5.0% to 3.5%, or 1.5 percentage points). In the context of a district serving 10,000 secondary students, each taking six classes a semester (60,000 grades each semester), this would translate into 900 fewer F's each semester, or 900 fewer "seats" needed in remedial or repeat courses[15] and, given the role of grades in pushing students to drop out of school,[16] could also result in fewer students leaving school. Particularly for those students historically underserved in schools–students of color and those from lower income families–reduction in grade depression opens doors for which they are qualified but may otherwise be closed to them simply because their teachers use traditional grading practices.

Implication #2: Improved, equitable grading practices lower rates of D's and F's assigned and reduce disparities in achievement. Course-aligned standardized test scores are the primary way to determine grade accuracy, and therefore help us to demonstrate the increased grade accuracy when teachers use improved grading. However, even if we remain skeptical of using standardized tests to determine grade accuracy, this study demonstrates a significant relationship between increased use of equitable grading practices and a decrease in D and F grades, and a narrowing of achievement disparities by race and income. Although, and perhaps because, improved grading practices are overlooked in both teacher preparation and in-service professional development, supporting teachers in this instructional area can, by itself, lower student failure rates and be an integral tool to reduce achievement and opportunity gaps.

Implication #3: Improved, equitable grading training for teachers is necessary, but not sufficient, to eliminate grade inflation and grade depression. Increasing the accuracy of grades depends on supporting teachers not just with improving their grading, but with other aspects of their professional work. This includes understanding course standards and content outcomes, knowing how to design an array of fair and valid assessments aligned with those outcomes, utilizing results of assessments to respond to student needs, and then to report those course outcomes accurately. Fortunately, because decisions about grading are implicated into nearly every instructional decision by a teacher ("Will I grade this task or activity, and if so, how?"), when teachers improve their grading practices, it can spur other instructional improvements;

three-quarters of teachers in this study reported that improving their grading practices generated significant improvements to their overall teaching.

Implication #4: Increased accuracy of teacher-assigned grades means that everyone–students, parents, counselors, college admissions officers, employers–can more confidently trust and rely on those grades. Decisions–ranging from which students should receive tutoring and which students should receive scholarships, to which students are placed into advanced or college-level courses and which students are eligible for the basketball team–will be based on more valid information about student performance and needs. College admissions officers and those making placement decisions at open-access and selective institutions, employers, school leaders and teachers, parents, and students can make more informed, objective, and fair decisions. Reduced grade inflation and reduced grade depression can generate greater public confidence in grades and strengthen the public's overall trust in teachers and schools.

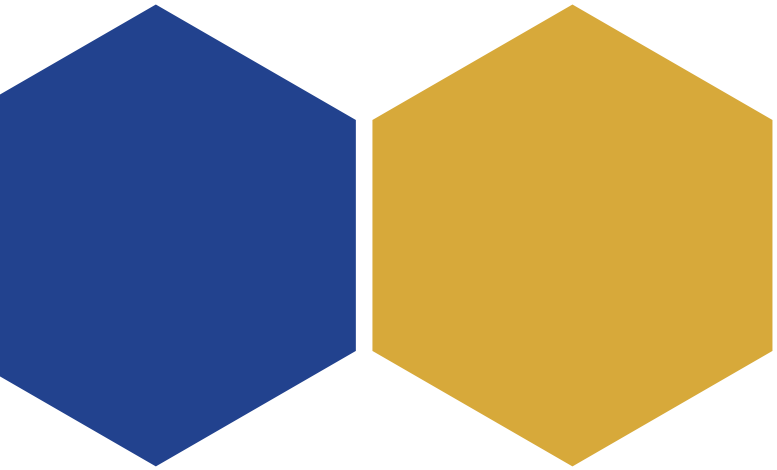
Recommendations

Recommendation #1: State and district leaders should prioritize addressing the "grading knowledge gap" among teachers as a necessary element of systemwide improvement to improve student achievement, reduce achievement gaps, and strengthen teacher professionalism and credibility, with benefits to student learning and district budgets. Teachers need to learn in both pre-service credentialing programs and in-service professional development to use grading practices that are more accurate, fair, and motivating. Prioritizing improved grading focuses a district or school community on some of the most important questions: What do we expect students to know and be able to do? What do we want to communicate about students' progress and achievement of these outcomes? Leaders should plan a multi-year effort that consists not only of significant training and support of teachers to improve their grading, but also the education of school board members, students and parents, updates to district and school policy and regulations, and adjustments to grading software and student information systems. This is a significant resource and financial investment, but aside from its benefits to student achievement, improving grading has financial returns: fewer failing grades means reduced remediation costs, and professional development in this "gap" in teacher skills and knowledge can reduce teacher attrition and save on recruitment, hiring, and training costs.

Recommendation #2: More research needs to be conducted that analyzes the impact of improved grading practices on both teachers and students. While this study was one of the largest conducted so far, what was not included was the significant qualitative data–testimonials, experiences, and stories–of educators who taught using these practices and of the students who learned with them. As one middle school teacher shared, "I used to think there was no other way to assign grades beyond the traditional method of classwork, homework, assessments, but now I think grading can be made more equitable and can more meaningfully reflect student understanding of content." Voices of teachers and students, especially of those students who often have received low grades in school, are crucially important to guide grading improvements.

Conclusion

Can students, like our student Olivia at the beginning of this paper, trust that their grades–one of the most permanent and consequential aspects of their K-12 education–are accurate? In other words, can a student, or anyone, trust a grade to tell the truth about that student's achievement? This study finds that when teachers use improved, more equitable and standards-based grading practices, the answer is yes. The consequences of these findings are far-reaching. School and district leaders and policymakers can make better decisions about system-wide improvements, college admissions officers and employers can make more informed decisions about specific students, and parents can know exactly where their children are in their learning. Most importantly, so can the children themselves. Improving teachers' grading is a powerful lever for lowering D and F rates, reducing grade inflation and grade depression, and making system wide improvements to our schools. What's more, these improvements to grading accuracy reduce achievement disparities for historically underserved students and can be a bulwark against the rising tide of mistrust in the important work of our K-12 schools and teachers.



Endnotes

1. Blad, Evie. "High Absenteeism Hits More Schools, Affecting Students With Strong Attendance, Too." *Education Week*, 12 October 2023, https://www.edweek.org/leadership/high-absenteeism-hits-more-schools-affecting-students-with-strong-attendance-too/2023/10
2. Natanson, Hannah. "Trust in teachers is plunging amid a culture war in education." *The Washington Post*, 6 September 2022, https://www.washingtonpost.com/education/2022/09/06/teachers-trust-history-lgbtq-culture-war/
3. This paper is not without bias, as it was conducted on behalf of an organization that provides services that improve teachers grading practices to be more accurate, bias-resistant, and intrinsically motivating. However, the grade-test consistency and grade distribution data was collected and evaluated by a third party evaluator, Elite Research Statistical Consulting, LLC to insulate its work from any interests the organization may have had in its outcomes.
4. Edge Research. *Parents 2019: As the Stakes Get Higher for Kids, Why Do We Lose Parents?* Learning Heroes, Sept. 2019, https://bealearninghero.org/wp-content/uploads/2019/09/Parents-2019-Research-.pdf
5. This report adds to the growing body of data showing a mismatch between teacher-assigned grades and tests. A recent analysis by The College Board found that the average GPA for high school students in English Language Arts and math courses increased from 2010 to 2022 while student test scores remained flat. Sanchez, Edgar, and Raeal Moore. *Grade Inflation Continues to Grow in the Past Decade*. ACT, Inc. 2022. https://www.act.org/content/dam/act/secured/documents/pdfs/Grade-Inflation-Continues-to-Grow-in-the-Past-Decade-Final-Accessible.pdf
6. Jones, Jeffrey M. "Confidence in U.S. Institutions Down; Average at New Low." *Gallup*, 5 July 2022, https://news.gallup.com/poll/394283/confidence-institutions-down-average-new-low.aspx
   Natanson, Hannah. "Trust in teachers is plunging amid a culture war in education." *The Washington Post*, 6 September 2022, https://www.washingtonpost.com/education/2022/09/06/teachers-trust-history-lgbtq-culture-war/
   Saad, Lydia. "Military Brass, Judges Among Professions at New Image Lows." *Gallup*, 12 January 2022, https://news.gallup.com/poll/388649/military-brass-judges-among-professions-new-image-lows.aspx

7. Institute of Education Sciences. "Percentage of undergraduates who reported ever taking a remedial course after high school graduation, percentage taking remedial courses in 2015–16, and percentage distribution of the number of remedial math and reading/writing courses taken in 2015–16, by control and level of institution and selected student characteristics: 2015–16—Continued" *U.S. Department of Education*, Table 1.7. https://nces.ed.gov/pubs2019/2019467.pdf
Institute of Education Sciences. "Remedial Coursetaking: Percentage of 2011–12 Beginning Postsecondary Students Who Took Any Remedial Courses and, Among Those Students, Average Number of Remedial Courses Taken and Passed, by Selected Characteristics: June 2017." *U.S. Department of Education*, Table 3. https://nces.ed.gov/pubs2020/2020501.pdf
8. Mowreader, Ashley. "Report: Grade Inflation Highlighted in Developmental Course Placement." *Inside Higher Ed*, 28 May 2024, https://www.insidehighered.com/news/student-success/academic-life/2024/05/28/colleges-place-students-remedial-courses-based-gpa
9. The 263 teachers, from 10 school districts in California, Virginia, and Iowa, taught a range of academic subjects in secondary grades (grades 6-12).
10. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. "Remedial Coursetaking." 2021, National Center for Education Statistics, 2009. https://files.eric.ed.gov/fulltext/ED532766.pdf
11. *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms* (2023)
12. It is worth noting that the number of Black students' grade-test pairings in the study was relatively small–1283 pairings out of 33,383 (less than 4%).
13. Institute of Education Sciences. "Teacher Attrition and Mobility Results From the 2021–22 Teacher Follow-up Survey to the National Teacher and Principal Survey." *U.S. Department of Education*, https://nces.ed.gov/pubs2024/2024039M.pdf;
Walker, Tim. "Survey: Alarming Number of Educators May Soon Leave the Profession." *National Education Association*, 1 Feb. 2022, https://www.nea.org/nea-today/all-news-articles/survey-alarming-number-educators-may-soon-leave-profession.
14. Administrators may be reluctant to engage teachers with improving grading, but importantly, only 4% of teachers reported that they were more likely to leave their school or district because of this professional development.
15. And if we include the 1.3 percentage point decrease in D's in this study– from 7.4% to 6.1%– that equates to a total of 1680 fewer D and F grades.
16. National Dropout Prevention Center. "Why Students Drop Out." *National Dropout Prevention Center*, 2024, https://dropoutprevention.org/resources/statistics/quick-facts/why-students-drop-out/.

Appendix: Methods for Grade-Test Score Consistency and Grade Distribution

School partners used a variety of different standardized testing methods with a range of different measurement types. An effort was made to standardize these tests into performance levels that could mimic the letter grades for the purpose of testing Grade-Test Consistency. Standardized tests provided by school partners were chosen on the basis of whether the tests had a substantial reach in the data (a large proportion of the students in the data had taken the test to maximize observations) and their ability to be converted into performance levels appropriately.

During the data preparation process, if standardized test scores were not accompanied by performance levels the standardized test scores were transformed into discrete 4-point scales using information from their respective testing manuals. If the performance levels were on a 5-point scale, simple transformations were performed to equate those scales to a 4-point scale. The standardized tests that already used a 4-point performance level were the Smarter Balanced Assessment Consortium (SBAC) and the STAR test. Tests that were converted to 4-point performance level using simple transformations were Advanced Placement (AP) tests (transformed from a 5-point scale) and the NWEA MAP (transformed from a 5-point scale). Tests that were converted to 4-point performance level by binning continuous scales into performance categories were the English Language Proficiency Assessments for California (ELPAC), the California Science Test (CAST) and the Virginia Standards of Learning (SOL). The Iowa Statewide Assessment of Student Progress (ISASP) had a 3-point scale and letter grades were converted to a 3-point to match the test performance levels and assess consistency.

School partners had test data for some students while some school partners had no test data for students (not state mandated standardized tests). In some instances, school partners sent more than one standardized test per student. For parsimony, it was decided that there would be one test per student per subject. For example, one student could have one ELA/Reading, one Math, and one Science test but a student could not have two Math tests to calculate the grade-test consistency score. Standardized tests were assessed to determine which test provided the most coverage within the dataset to determine which standardized test maximized the number of observations in the data. If a student had only one test type available per subject, that test would be used for that student's observation. If there were two or more standardized tests competing for the same student's subject, to narrow down which test was retained, the standardized test that maximized the number of observations was chosen as the preferential test. In some instances, there were students that had different standardized test types depending on the subject (e.g. AP for Science but ELPAC for ELA).

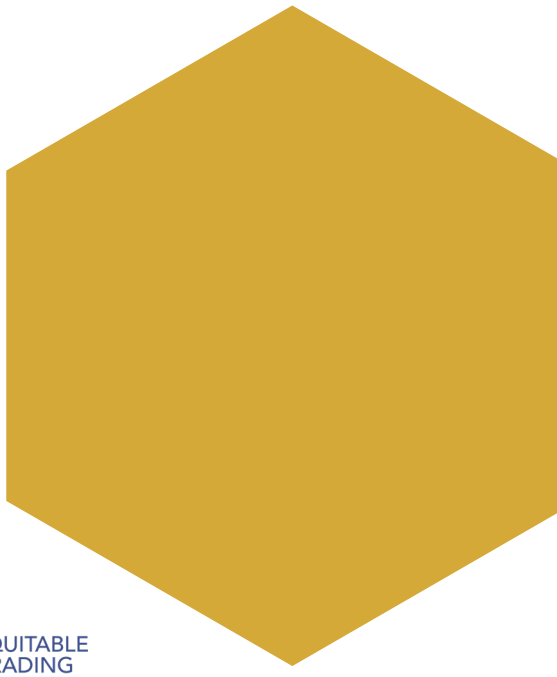
For the Grade-Test Consistency assessment, course letter grades were transformed from the typical five-letter grade assessment into a four-letter grade assessment by collapsing D and F letter grades into the same group. The reason

for collapsing these letter grades was due to the extremely low counts in the F group which could cause issues with estimation and because many standardized tests had performance levels at four levels and not five. Once grades and tests were standardized at the four performance levels and the four course letter grades, these indicators were subtracted from each other to assess the level of grade-test consistency. If this grade-test consistency difference score was 0 (meaning complete consistency between tests and grades) or ±1 (meaning some variation between tests and grades) they were coded as a 1 indicating adequate grade-test consistency. If the grade-test consistency difference score was ±2 or more, those observations were coded a 0 indicating inadequate grade-test consistency. This binary coding of grade-test consistency was used in the primary analysis. The five-level grade indicator was used in the primary analysis to assess how grade distribution changed. The Grade-Test Consistency analysis comprised teachers who taught in both analysis years, whereas the Grade Distribution analysis was exclusively teachers who taught in both years and also taught the same course in both analysis years. Additionally, during the analysis process certain samples reduced further in size if the observations were missing the demographic variables included in the models. Multiple school partner data was then combined into a single aggregate dataset to assess the overall EGP program impact.

For the Grade-Test Consistency analysis, multilevel logistic regressions were conducted using the binary coding of grade-test consistency as the outcome of interest and interacting time (pretest and posttest) and whether the students teachers were Basic EGP (teachers with 1 year of EGP PD), Advanced EGP (teachers with 2 years of EGP PD), or non-EGP (if applicable – some schools did not provide control students) while controlling for student demographics (gender, race/ethnicity, free or reduced price lunch, English-language learner, if the student had an IEP, and subject type). In addition, subgroup analysis looking at gender, ethnicity/race, and FRPL differences were also conducted. Marginal probabilities of these analyses were estimated to determine, holding all other effects constant, what was the likelihood of grade-test consistency from pre to post given these subgroup differences.

For the Grade-Distribution analysis, multilevel ordinal logistic regressions were conducted using the five group letter-grade coding (A through F) of grade-distribution as the outcome of interest and interacting time (pretest and posttest) and whether the students teachers were Basic EGP (teachers with 1 year of EGP PD), Advanced EGP (teachers with 2 years of EGP PD), or non-EGP (if applicable) while controlling for student demographics (gender, race/ethnicity, free or reduced price lunch, English-language learner, if the student had an IEP, and subject type). In addition, subgroup analysis looking at gender, ethnicity/race, and FRPL differences were also conducted. Marginal probabilities of these analyses were estimated to determine, holding all other effects constant, what was the likelihood of grade-distribution from pre to post given these subgroup differences.

EGP developed its own dosage indicator that measured the extent to which a teacher was implementing equitable grading practices in their classrooms. This indicator was based on a teacher self-report survey which asked the teachers twelve questions that measured the frequency of implementing various equitable grading practices that they learned in their professional development training workshops. These questions were averaged together into a composite indicator of dosage, merged into the student data, and was used in the analysis to determine if there was a difference in the level of dosage on grade test consistency and grade distribution outcomes.



EQUITABLE GRADING PROJECT

info@equitablegradingproject.org