Course Grades as a Signal of Student Achievement: Evidence on Grade Inflation Before and After COVID-19

Dan Goldhaber & Maia Goodman Young
CALDER Research Brief No. 35 November 2023

CALDER
Center for Analysis of Longitudinal Data in Education Research
at the American Institutes for Research®

"America's schoolchildren face a daunting task to recover what they've lost. The least schools can do is tell the truth about how far they have to go." – Washington Post, 2/22/2023

"Despite the importance of grade inflation and the widespread reports of it, there has been little systematic research exploring changes in grading standards—which would include grade inflation —in U.S. high schools" - Koretz & Berends (2001, p. iii)

"No rules are etched in stone that set eternal grading standards." - Blount (1997, p. 330)

Introduction

While grading has been a topic of research for well over a century,[1] teacher *grading standards* are receiving increased attention—and with good reason.[2] There is widespread speculation (e.g., Johnson, 2021; Klinger et al., 2022; Mathews, 2022; Walker, 2021) and some evidence (e.g., Sanchez & Moore, 2022, Sanchez, 2023) that grading standards have changed over the course of the pandemic, making higher grades relatively easier to achieve and less reflective of objective measures of learning. It is possible—even likely—that shifting grading standards give parents, guardians, and students a confusing or inaccurate picture of what students know and can do, especially considering pandemic-related learning losses (Dorn et al., 2021; Goldhaber et al., 2023; Kuhfeld et al., 2022). Indeed, recent pieces in the *Los Angeles Times*, *New York Times* and *The Associated Press* raise this alarm, observing that many parents are left in the dark by the lack of information from their children's schools. Public opinion surveys point to a discrepancy between what parents believe about their student's level of achievement, i.e., that students have recovered academically, and what test results like NAEP suggest about their achievement (Esquivel, 2022; Kane & Reardon, 2023; Vázquez Tonnes, 2023).

Aside from issues of communication about educational achievement, there is some empirical evidence that a loosening of grading standards could be detrimental to students' learning.[3] Studies have linked higher grading standards to greater student achievement later in a student's schooling (Betts & Grogger, 2003; Bonesrønning, 2004; Figlio & Lucas, 2004; Gershenson, 2022; Mozenter, 2019).[4] In a recent study of changing grading policies in North Carolina, researchers found that more lenient grading policies led to more absenteeism and lower test scores for students at the lower end of the testing distribution (Bowden et al., 2023). At the same time, there are arguments supporting changes in grading standards during the pandemic when remote instruction made it difficult for teachers to assess students. There was also fear that low grades might discourage students as they reengage with schooling post-pandemic (Feldman and Reeves, 2020).

Despite considerable theoretical work about grading standards and some evidence of changes in grading over the pandemic, we have limited research on the extent to which eased grading standards continued post-pandemic, as teachers and students returned to normal schooling. In this research brief, we use administrative data on student grades from Washington state to assess whether grading standards have returned to pre-pandemic levels.

We analyze nearly a decade of middle and high school course grades in Washington state before and after the pandemic to assess how grades changed over time within courses. We find that average grades in math, English, and science courses rose slightly in each subject in the decade before the pandemic. These grade increases corresponded with a period of generally increased test achievement. During the pandemic, explicit guidance from the state about easing grading standards led, not unexpectedly, to large increases in grades in each subject. For instance, the average GPA in math jumped 0.34 points from 2.36 to 2.70—roughly the difference between a C+ and B- — between the 2018-19 school year and the 2020-21 school year (after having only increased 0.11 points between 2011 and 2019). Following the pandemic, grades returned to pre-pandemic averages in most subjects. But test achievement is far below its pre-pandemic levels—including in Washington State (e.g. Bazzazz 2022, 2023), hence we might expect a greater divergence between the grades students receive and their standardized test scores. Indeed, we see such a divergence between math grades and test scores following the pandemic.

Grading Before and After the Pandemic

Grading students on their work serves multiple purposes: it communicates achievement to students and their families; encourages effort, self-esteem, and self-reflection; determines appropriate educational interventions; and can inform the evaluation of educational programs (Blount, 1997; Brookhart et al., 2016, 1993; Guskey & Bailey, 2001; Reeves, 2011). Grades are also, of course, used across institutions to judge merit, including whether students are ready for particular courses of study and admission to colleges and universities (Brookhart, 2011; Castro et al., 2020). But grades are often a "hodgepodge measure," based on school or teacher norms and comprised of cognitive and noncognitive measures (Brookhart, 1991; Brookhart, et al., 2016; Bowers, 2011).[5]

Despite the importance of grades as a measure of students' success and potential, there is little in the way of consistent guidance for how grading should be done or the degree to which grades should reflect objective measures of learning. This uncertainty is particularly problematic in cases where schools must make decisions that adjust grading standards, as during the COVID-19 pandemic when districts adopted policies that made it less likely that students would receive low grades (Malkus et al., 2020; Reich et al., 2020; Townsley, 2020). Such adjustments have the potential to change the relationship between grades and achievement in academic subjects, and thus confuse the signal grades send to stakeholders (Opalka & Talkington, 2020; Vahle et al., 2023).

There is abundant evidence that grading policies changed at the outset of the pandemic (Malkus et al., 2020; Reich et al., 2020; Townsley, 2020). For example, many states adopted policies based on the principle of "do no harm" (Opalka & Talkington, 2020). These policies varied within states by district, from giving nearly all students "As" (Bazzaz & Long, 2020), to only assigning credit/no credit grades for the spring of 2020, to mandating that no student's grade should drop from what it was when schools shut down (Reich et al., 2020; Townsley, 2020).

These emergency grading policies generated much debate in the education community. While many argued for more relaxed grading policies (which would equate to more inflated grades and easier grading standards), others argued that these policies would do harm to some students by eliminating a major source of motivation to stay in school.[6] Hamilton et al. (2020) reported that around 80% of teachers surveyed reported requiring students to complete work, but only one-third were assigning letter grades for the work. Secondary teachers reported being more likely to grade, with 50% assigning grades and only 14% monitoring for completion. Malkus et al. (2020) found similar patterns in an analysis of school websites: 35% of schools in the sample graded based on completion while 32% graded on performance. Only 12% of the schools in the sample explicitly stated that new work would not be graded (Malkus et al., 2020).

These changes to grading likely led to grade inflation, at least temporarily. While there is no standard definition of grade inflation (see Tyner & Gershenson, 2020), researchers tend to study it by comparing student's course grades (Frey & Birnholz, 2020; Gerhenson, 2018; Lekholm & Cliffordson, 2009) or student's academic GPA (Camara et al., 2003; Woodruff & Ziomek, 2004) with standardized test scores to determine if grades and scores diverge.[7] Recent studies using data from the ACT suggest the potential impact of COVID-era grade inflation (Sanchez, 2023; Sanchez & Moore, 2022). These studies are informative but also have limitations: they only cover students who are likely bound for college, and the nature of the data makes it difficult to pin down how the relationship between grades and test scores may have changed due to the pandemic.

There is anecdotal evidence suggesting that some educators support maintaining pandemic-era shifts in grading standards (e.g., Simonetti, 2020; Vasquez, 2021). In meetings with students, teachers, principals, district administrators, and parents, for example, Reich and Mehta (2021) identified "eliminating the system of averaging grades which heavily penalizes students with zeros for incomplete assignments" as a practice that was worth growing (p. 14). Penuel (2021) acknowledged that some educators "have been just as eager to let go of old educational practices" and maintain pandemic-related changes to grading (p. 54). Systematic evidence about whether pandemic-related grade inflation has endured, however, is limited.

Pandemic-era Grading in Washington state

In Washington state, grading guidance evolved over the early part of the pandemic. At the onset of the pandemic in March 2020, the state's Office of Superintendent of Public Instruction (OSPI) offered suggestions to school districts, such as "to honor student work, consider adoption of pass/no credit grading, competency-based credit, or other grading methods" (Reykdal, 2020a). Subsequent guidance from OPSI was more directive. On April 21, 2020, for example, OSPI instructed teachers not to assign grades of "pass," "fail," or "no credit" to students in grades 9-12 or to middle school students taking high school credit bearing classes such as Algebra 1. Teachers were allowed to assign an "incomplete" to students who could not engage in schooling after schools closed on March 17. The state also mandated that districts give students with "incompletes" opportunities to "reengage in the learning standards" through later coursework (Miller et al., 2020, p. iii).[8] No student would receive a lower grade than they had when schools around the state shut down in March (*Student Learning & Grading Guidance*, 2020).

Guidance from OSPI changed again in the fall of the 2020-21 school year, when some districts resumed in-person or hybrid schooling. During this period, teachers were allowed to assign failing grades, such as "F," "fail," or "no credit" (Reykdal, 2020b). Districts responded with a range of grading policies. In Seattle Public Schools, for example, the district instructed teachers to give students grades only between an A and C-, or an incomplete (Seattle Public Schools Board of Directors, 2020). In other districts, teachers went back to an A through F grading scale (e.g., Pasco Board of Directors, 2020).

Washington's grading policies during the early pandemic were not unusual. The call to "do no harm" appeared in several state and district grading policies during this period, as states, like Washington, urged districts to prioritize student physical and emotional safety. In some places, districts engaged families and students to determine the best grading policy to implement in a crisis (Malkus et al., 2020, Reich et al., 2020, Townsley & Kunnath, 2022). Whether any of these changes to grading standards persisted in the years since the pandemic is unclear.

To investigate how grading standards have (or have not) changed since the pandemic, we use student-level data from the Comprehensive Education Data and Research System (CEDARS), which includes test scores and high school credit-bearing grades for students from school year 2010-11 to 2021-22 in Washington state. We observe high school course grades for most of the students in the sample and also observe some course grades for middle school students (grades 7 and 8) who took credit-bearing classes, such as Algebra 1 and Geometry, in middle school. In the CEDARS dataset, courses are identified by state course identification number, state course name, and local course name. Using these labels, we can identify the courses of interest.[9] In order to calculate average subject GPAs, we converted grades to a 4.0 scale.[10]

Table 1 shows the two student samples we use. The first sample—the "grade sample"—spans from the 2010-11 school year to the 2021-22 school year and contains students who took a math, English, or science class in 10th or 11th grade and the students who took Algebra 1 or Geometry in 7th or 8th grade. We use this sample to describe how the shares of grades and average GPAs have shifted over time within content areas. The second sample—the "grade-test score sample"—includes students for whom we observe both course grades and test scores. Due to changes in state standardized testing regimes, we use test score data only from the 2015-16[11] school year to the 2021-22 school year for the Smarter Balanced Assessment (SBA) in math (grades 7, 8, 10, and 11 for Algebra 1, grades 8, 10, and 11 for Geometry, and grades 10 and 11 for Algebra 2) and scores from 2014-15 to 2021-22 in English (grades 10 and 11).[12] We do not observe both grades and tests for students who take courses in the 9th grade because there is not a statewide assessment in the 9th grade. We also use data from only the 2017-18 school year to the 2021-22 school year for 11th grade students taking the Washington Comprehensive Assessment of Science (WCAS). Because there was no state testing in spring 2020 or spring 2021, we do not have test scores for those school years.[13] We use

Table 1. Number of Students by Subject, Year, and Grade Level

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020F	2020S	2021	2022
Algebra 1													
7th grade Algebra 1	1,348	1,434	1,183	729	1,054	1,163	986	877	756	4,674	5,404	5,467	4,931
as a % of 7th graders	1.72	1.82	1.49	0.92	1.34	1.46	1.21	1.07	0.90	5.33	6.16	6.36	5.94
8th grade Algebra 1	6,756	5,352	5,866	3,349	4,376	4,696	4,829	3,142	2,826	13,686	16,177	16,859	17,533
as a % of 8th graders	8.61	6.78	7.40	4.19	5.44	5.91	5.96	3.84	3.45	16.08	19.01	19.56	20.60
10th grade Algebra 1	18,098	15,946	16,167	15,708	14,507	13,526	12,231	12,145	11,738	10,094	9,932	9,879	13,067
as a % of 10th graders	22.07	19.85	20.03	19.34	17.71	16.32	14.65	14.79	14.17	11.95	11.76	11.78	15.46
11th grade Algebra 1	10,323	9,745	9,042	8,941	8,774	7,452	5,793	5,970	7,322	4,930	5,080	5,182	5,375
as a % of 11th graders	12.99	12.24	11.52	11.29	10.83	9.14	6.96	7.17	9.01	5.95	6.13	6.23	6.51
Geometry													
8th grade Geometry	1,163	1,312	782	745	1,143	1,010	1,079	836	845	4,928	5,719	5,570	5,570
as a % of 8th graders	1.48	1.66	0.99	0.93	1.42	1.27	1.33	1.02	1.03	5.79	6.72	6.46	6.54
10th grade Geometry	30,876	33,626	34,712	33,414	35,610	36,987	38,941	39,026	41,339	36,298	39,084	39,922	39,922
as a % of 10th graders	37.66	41.87	43.02	41.15	43.47	44.61	46.63	47.51	49.91	42.97	46.27	47.61	47.23
11th grade Geometry	10,234	11,646	11,185	10,719	10,984	10,529	9,697	8,576	9,295	6,120	7,039	8,664	8,664
as a % of 11th graders	12.88	14.63	14.25	13.54	13.56	12.91	11.65	10.31	11.44	7.38	8.49	10.42	10.49
Algebra 2													
10th grade Algebra 2	16,808	20,452	20,708	19,693	19,421	20,920	20,546	20,975	20,571	18,250	19,404	20,961	19,946
as a % of 10th graders	20.50	25.46	25.66	24.25	23.71	25.23	24.60	25.54	24.83	21.61	22.97	25.00	23.60
11th grade Algebra 2	19,189	24,274	25,387	23,950	24,149	26,073	25,874	26,291	27,384	25,543	25,530	28,388	28,731
as a % of 11th graders	24.15	30.48	32.35	30.25	29.80	31.97	31.07	31.59	33.71	30.82	30.80	34.14	34.78
English													
10th grade English	65,566	65,796	67,275	65,355	68,921	70,669	70,748	81,985	82,703	73,144	76,304	79,089	82,037
as a % of 10th graders	79.97	81.92	83.37	80.48	84.14	85.24	84.72	99.81	99.84	86.59	90.33	94.32	97.05
11th grade English	48,244	50,818	51,452	51,728	55,480	56,160	47,402	51,339	67,172	57,624	59,957	64,407	68,279
as a % of 11th graders	60.71	63.82	65.56	65.34	68.47	68.86	56.93	61.70	82.68	69.52	72.33	77.45	82.67
Science													
10th grade Science	65,770	66,573	69,379	67,989	70,888	73,860	73,154	75,021	77,237	66,709	68,910	72,971	75,515
as a % of 10th graders	80.22	82.89	85.98	83.73	86.54	89.09	87.60	91.34	93.25	78.97	81.58	87.02	89.33
11th grade Science	39,741	40,956	40,564	39,890	40,891	40,553	32,670	28,968	41,714	37,988	38,846	42,500	44,960
as a % of 11th graders	50.01	51.43	51.68	50.39	50.47	49.73	39.24	34.81	51.34	45.83	46.86	51.11	54.43

Notes: (1) 2011 refers to the 2010-2011 school year (other years follow the same pattern) (2) Because of the change in grading policy in the spring of 2020, we observe students separately in for both the fall of 2019-20 and the spring of 2019-20. (3) We only include in our sample students who have a letter grade for the course of interest; thus, even if every student took an English class in 10th grade, we would still only observe a percentage of those students because some of these students would receive grades like “credit” or “incomplete.” 4) Enrollment data from the National Center for Education Statistics.

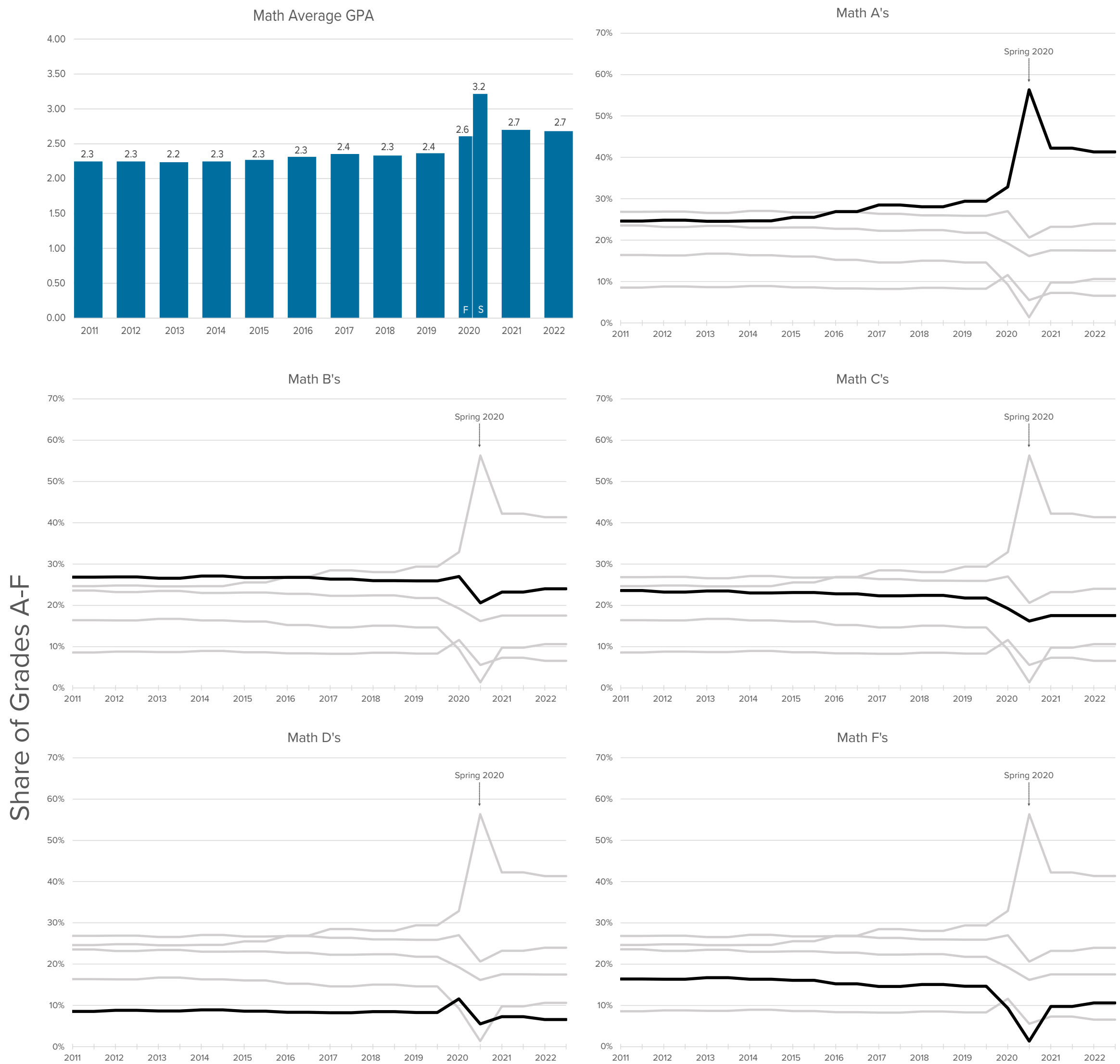
all the scaled scores across years and normalize them within grade and subject to have a mean of 0 and a standard deviation of 1. Note that this captures the full range of student performance in the sample, and that we purposely *do not* standardize within year so that changes in the entire distribution across years affects the year-to-year estimated relationship between grades and tests.

Each cell in Table 1 shows the number of students that are in the grade sample. In years where we observe test scores, around 90% of math and English students who have grades also have test scores. In science, around 70% of students who have grades also have test scores.[14] Below the number of students with grades, we include the percentage of students in that grade level who have grades in that subject area. For instance, in 2011, under 2% of 7th grade students took Algebra 1 (1.72%). In the same year, just over 22% of 10th grade students took Algebra 1.[15]

Evolution in Grades and Test Results Over Time

As others have noted (e.g., Sanchez, 2023), grade point averages have fluctuated over the past decade. We find similar trends in our data, particularly in math (see **Figures 1a-1c**). Figures 1a-1c provide information about course grade for cross-sections of students in selected courses over time. The chart in the upper left of the figures depicts the average grade students received in the subject by year. Again, we split the 2019-20 school year into

Figure 1a. Distribution of Math Grades A-F, 2011-2022



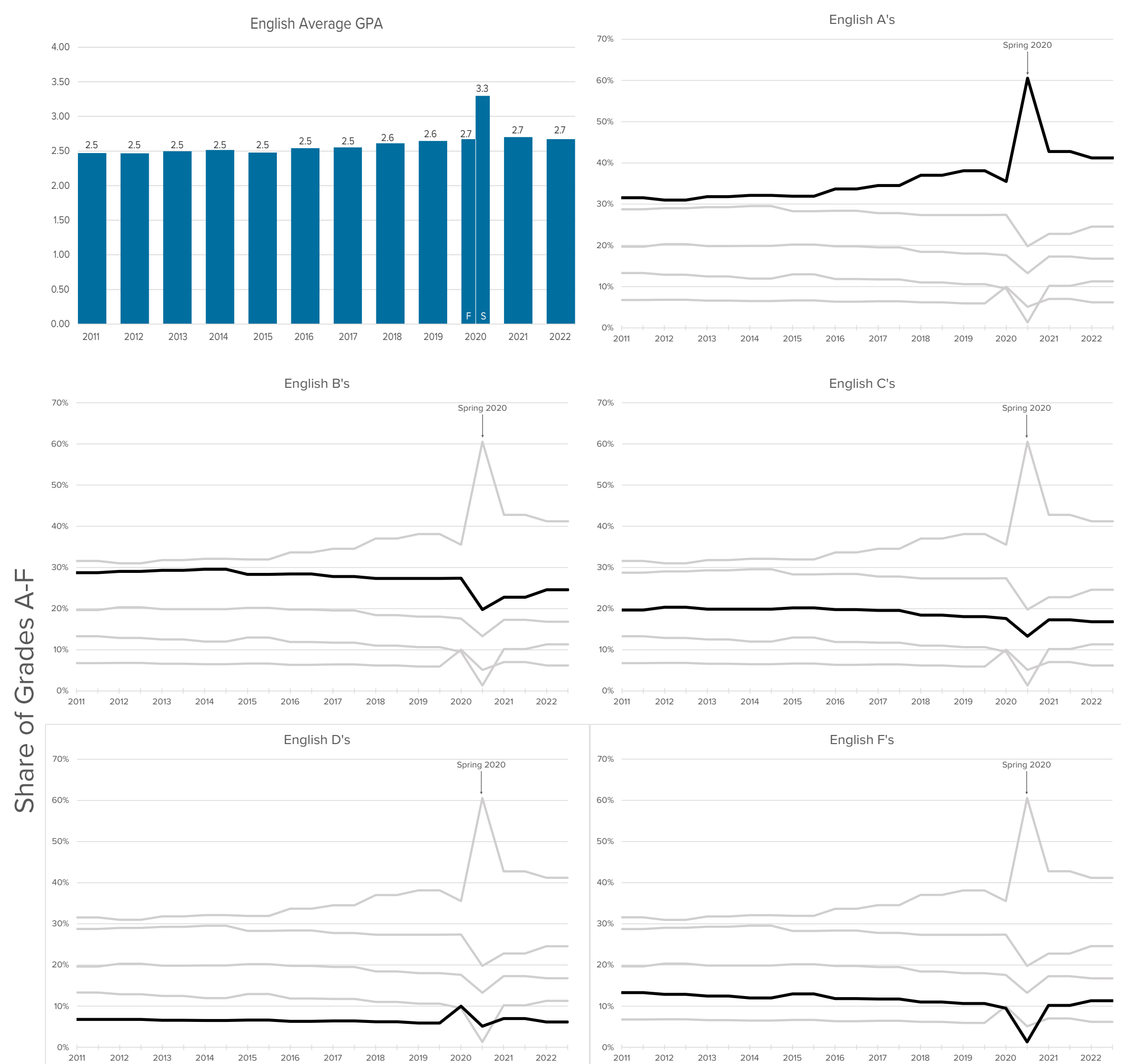
Notes Figures 1a-1-c: (1) Because districts differ in the number of grades students receive (e.g., on school may give students one grade per trimester while another may give students a single grade for the entire year), we averaged grades to the student, subject, and school year level. (2) As with our analytic sample, these graphs contain the grades for tested courses where students also receive high school credit. Thus, we include grades from 7th , 8th, 10th and 11th grade math classes including Algebra 1, Geometry, and Algebra 2; 10th and 11th grade English; and 11th grade science courses including biology, chemistry, and physics.

fall and spring to indicate the onset of the pandemic. The line charts indicate the proportion of students receiving each letter grade, with the heavy line indicating the share of students who received that particular grade.[16]

Prior to the pandemic, the figures show a slight uptick in math and English grades. For instance, in math, average GPAs increased 0.1 points between 2010-11 and 2018-19. In English and science, GPAs increased about the same amount over the same period. However, math GPAs increased in the fall of 2019-20, before any school closures or grading guidance from the state. Grades in math jumped 0.2 points between the 2018-19 school year and the fall of the 2019-20 school year—more than the total increase in math GPAs in the previous eight years.[17]

Consistent with Washington state guidance, Figures 1a-1c show that almost no students received an F grade in the spring of 2020. The share of F grades dropped from 9.5% to 1.3% in English courses, 9.3% to 1.4% in math courses, and 8.6% to 1.1% in science between the fall and spring semesters of 2020. The distribution of grades higher than F mostly increased for A grades, with the share of A's jumping from 32.9% to 56.3% in math, from 35.5% to 60.6% in English, and from 32.5% to 59.0% in science. The average GPA in math jumped from 2.6 to 3.2, in English it jumped from 2.7 to 3.3, and in science it jumped from 2.6 to 3.3. The figures also suggest that English and science grades largely returned to pre-pandemic levels by 2021-22, but math grades did not. Indeed,

Figure 1b. Distribution of English Grades A-F, 2011-2022



the math GPA in 2021-22 was 2.7, 0.4 points higher than it was in 2018-19.

To better understand what these shifts in grading might mean, we perform some simple regressions to descriptively assess the extent to which the relationship between grades and test scores has changed over time. Specifically, we regress individual student test scores by subject on grades in that subject, the year the test was taken, the grade level (7th, 8th, etc.) in which the subject and test were taken, and interactions between the grade level and year and the grade level and grade received.[18] We do this to account for the fact that the grade level at which students take a subject such as Algebra 1 may be correlated with achievement in that subject, and there may be shifts over time in the likelihood that students take a particular subject in a particular grade.[19]

In **Figure 2** on the next page, we report a student's predicted place in the test distribution based on the letter grade received.[20] As is apparent from visual inspection, a student's predicted place in the test distribution drops for all math subjects between the first year of observed test scores (2015-16) and the first year of testing after the pandemic (2021-22). For instance, a student who got an "A" in Algebra 1 was predicted to be in the 73rd percentile of the test distribution in 2015-16, the 68th percentile in 2018-19, and the 58th percentile in 2021-22. In Algebra 2, a student receiving an "A" was predicted to be in the 64th percentile in 2015-16, the 58th percentile in 2018-19, and the 54th percentile in 2021-22.

Figure 1c. Distribution of Science Grades A-F, 2011-2022

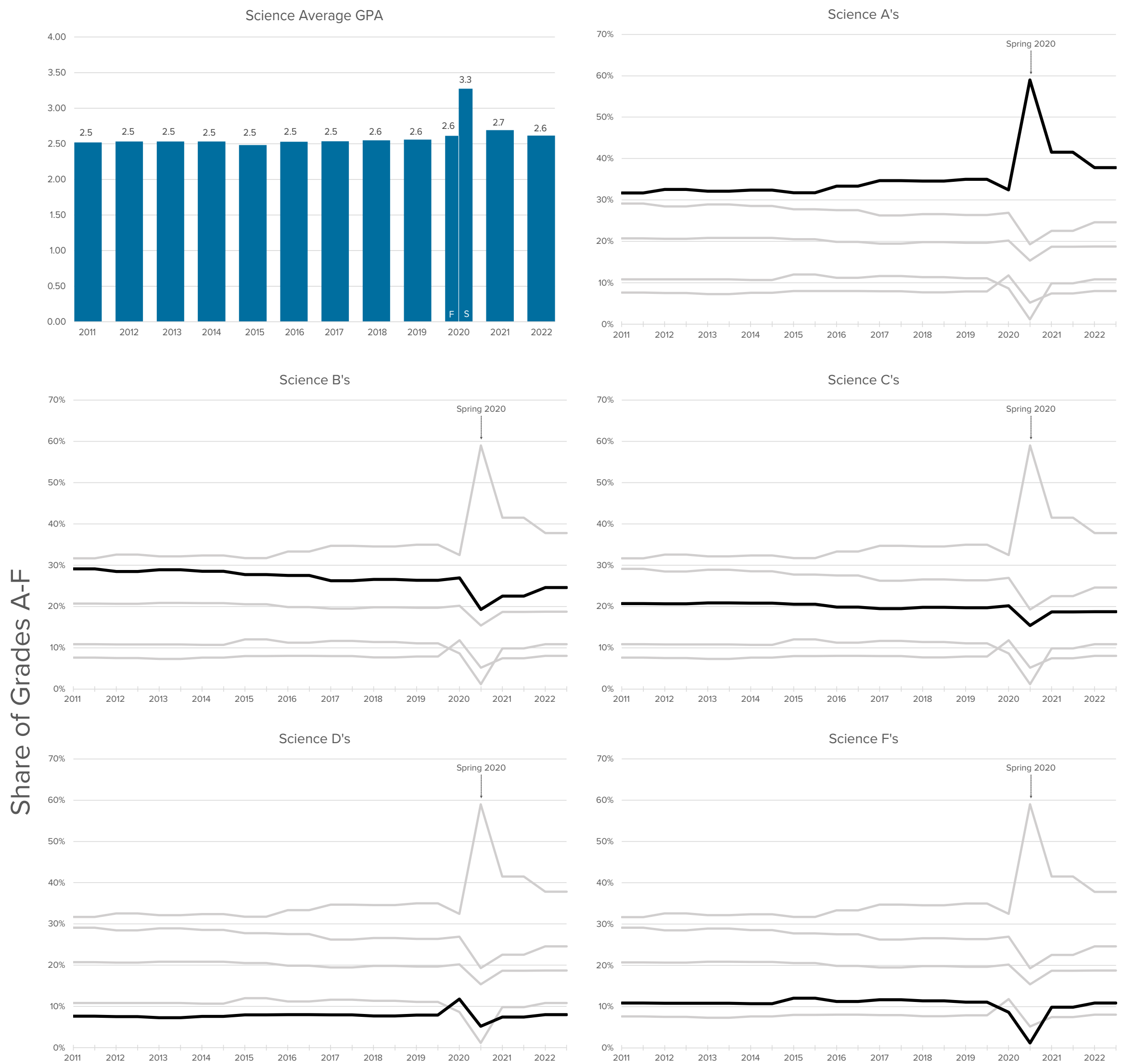
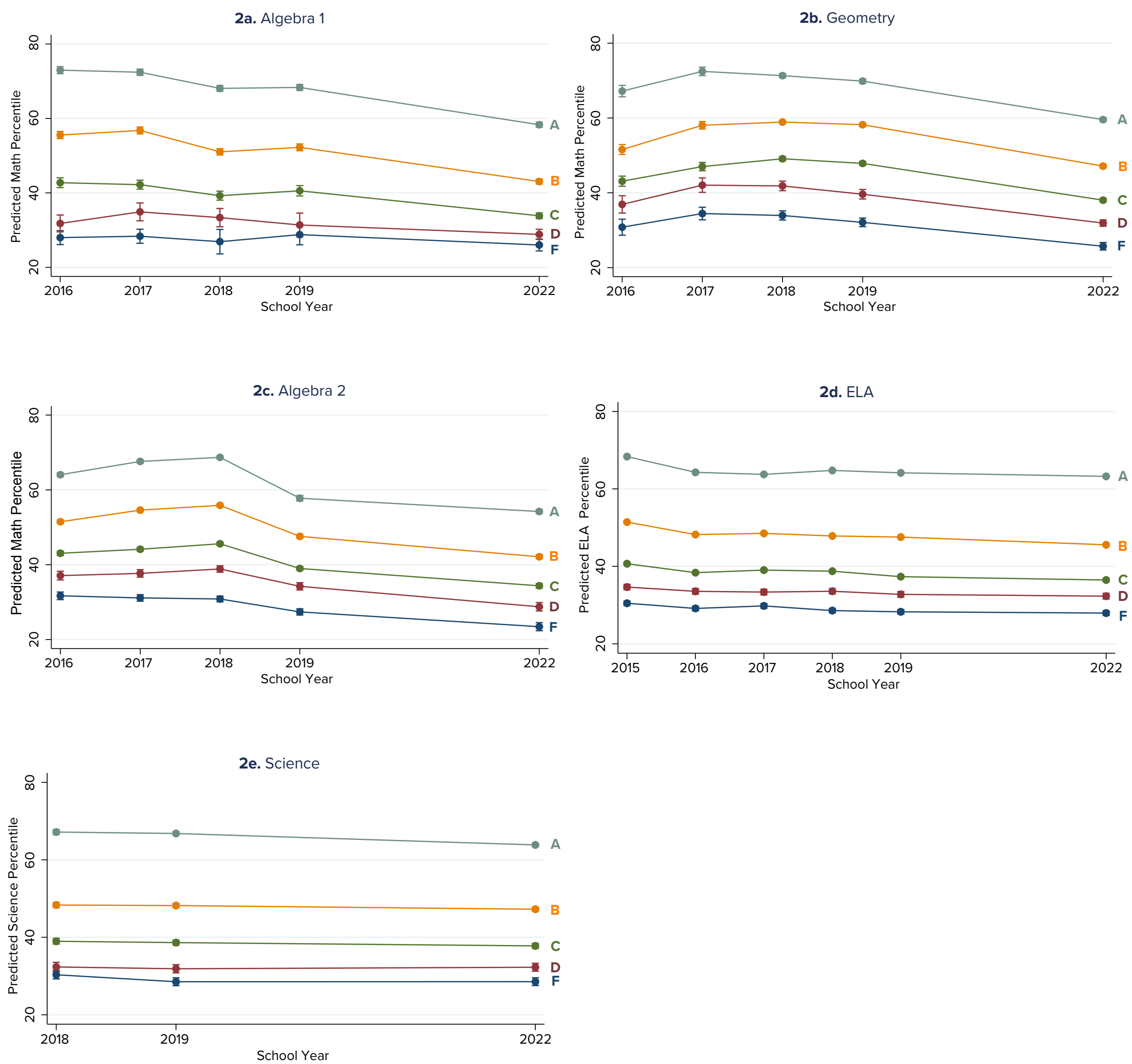


Figure 2. Predicted Place in Test Distribution by Letter Grade across Courses



However, these shifts are not uniform across all letter grades for all subjects. For instance, when we focus on "F" grades in Algebra 1, we observe that students receiving an "F" were predicted to fall roughly in the same spot in the test distribution—somewhere between the 27th and 29th percentile—across all the years. Whereas, in Algebra 2, a student receiving an "F" in 2015-16 was predicted to be in the 32nd percentile, while a student receiving an "F" in 2021-22 was predicted to be in the 23rd percentile.

In English, the shifts in grade standards are less dramatic, yet still noticeable with the predicted test scores by grade declining gradually between 2014-15 and 2018-19, and then staying roughly even between 2018-19 and 2021-22. Similarly, in science, we observe only small shifts.

Another trend is notable in Algebra 1. In each year from 2015-16, there is a narrowing of the range of predicted test scores associated with significantly different letter grades received. For instance, in 2015-16 a student receiving an "A" is predicted to fall in the 73rd percentile and a student receiving an "F" is predicted to fall in the 28th percentile—a difference of 45 percentile points. In 2021-22 that difference falls to 32 percentile points. This trend is not consistently true for the other subjects. In Algebra 2 we see a widening of this range from 32 percentile points in 2015-16 to 38 percentile points in in 2017-18 before the range narrows back to 31 percentile points in 2021-22. And in Geometry, Science, and English, the range stays relatively stable across all years observed.

Discussion and Conclusions

In this brief we document significant changes in grades over time, especially with the onset of the pandemic. We find modest increases in student grades in the decade before the pandemic that accelerated (consistent with state guidance) after the pandemic's onset. Grades decreased in the year after the pandemic, but still tended to be higher, especially in math, than they were before the pandemic. We also see evidence, again especially in math, that the relationship between grades and test scores has diminished over time. These results are descriptive and do not illustrate the degree to which grading standards might vary across contexts, such as school system type, pandemic-related closures, or across student subgroups and test achievement level. Such distinctions are the subject of our ongoing work.

It is not clear how to view these findings. For instance, we recognize that schools are still working on reengaging students and regularizing schooling. Chronic absenteeism, which emerged with remote schooling is still a troubling issue in schools today (Leonhardt, 2023). A too fast return to pre-pandemic grading standards could serve to alienate some of the very students most in need of reengagement with school. Moreover, it is important to recognize that we are using only one metric to judge grading standards, the relationship between grades and test scores. It would be good to see evidence about how grades comport with other measures of student engagement and performance, such as attendance, and how that has changed over time.

But others may be more concerned about the misalignment between grading and student achievement, especially in math. Algebra 1—the course for which we noted the greatest weakening in the relationship between test scores and grades—is seen as a gateway to more advanced math concepts (Snipes & Finkelstein, 2015). It is arguably important for students, families, and schools to have accurate signals about students' achievement levels in this foundational course. Schools use grades in classes such as Algebra 1 to determine whether students need extra support, remediation, or even if they must repeat a course before moving on. If this signal is no longer accurately conveying a student's level of achievement, school systems risk under-supporting students who need help.

Likewise, families and students use grades as a signal of how a student is doing in school; the expectation is that if a student is having academic trouble, that trouble will show up in their grades. Decisions such as whether to enroll a child in after school tutoring or summer school may rest on a belief that the grades on a report card accurately reflect a student's levels of achievement. As we noted above, many parents are under the impression that their children are not suffering from learning loss due to the pandemic; however, test scores indicate otherwise. It is possible that without a grade that signals trouble, parents may not choose to get needed extra academic support.

It is also important to consider that grades may be seen as representative of system level expectations of students (Lemov, 2023). As such, a relaxation of the standards goes against some of what we have learned about the importance of having high expectations of students. Making it easier for students to receive high grades may appear more equitable in the short run (in that, by definition, students will all receive higher grades), but the longer-term effect of such leniency has differential outcomes for higher- and lower-achieving students: benefiting the former and further disadvantaging the latter (Bowden et al., 2023).

While—as we noted in the introduction—there is a long history of research on grade inflation, there is a more limited examination of how grading standards have shifted over time. There has also been limited investigation of the forces that might push grading standards higher or lower. While there is ample anecdotal evidence that pressure from students, parents, or administrators can influence a teacher's grading standards, we don't yet have much evidence about the weight of these forces or how state or local policies may impact standards. This is an important area of further exploration going forward.

References

A nation at risk: The imperative for educational reform. (1983). National Commission on Excellence in Education.

Babcock, P. (2010). Real costs of nominal grade inflation? New evidence from student course evaluations. *Economic Inquiry*, *48*(4), 983–996. https://doi.org/10.1111/j.1465-7295.2009.00245.x

Bazzaz, D. (2022, January 18). Washington students' test scores drop significantly in first exams since pandemic began. *The Seattle Times*.

Bazzaz, D. (2023, September 14). New report: WA student test scores inching up in math and reading. *The Seattle Times*. https://www.seattletimes.com/education-lab/wa-student-test-scores-in-math-and-reading-inch-up-newest-release-shows/

Bazzaz, D., & Long, K. (2020, April 20). Seattle will give high-school students A's or incompletes; colleges urged to adopt generous grading. *The Seattle Times*.

Blount, H. P. (1997). The keepers of numbers: Teachers' perspectives on grades. *The Educational Forum*, *61*(4), 329–334. https://doi.org/10.1080/00131729709335278

Bowden, A. B., Rodriguez, V., & Weingarten, Z. (2023). *The unintended consequences of academic leniency* (EdWorkingPaper 23–836). Annenberg Institute at Brown University. https://edworkingpapers.com/ai23-836

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

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, *30*(2), 123–142.

Brookhart, S. M. (2011). *Grading and learning: Practices that support student achievement*. Solution Tree.

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

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E. A. (2003). *Whose grades are inflated?* (Research Report No. 2003–4). College Board.

Castro, M., Choi, L., Knudson, J., & O'Day, J. (2020). *Grading policy in the time of Covid-19: Considerations and implications for equity* [Policy and Practice Brief]. California Collaborative on District Reform.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2015). The aftermath of accelerating algebra: Evidence from district policy initiatives. *The Journal of Human Resources*, *50*(1), 159–188.

Dorn, E., Hancock, B., & Sarakatsannis, J. (2021, July 27). *COVID-19 and education: The lingering effects of unfinished learning*. McKinsey & Company. https://www.mckinsey.com/industries/education/our-insights/covid-19-and-educationthe-lingering-effects-of-unfinished-learning

Esquivel, P. (2022, December 22). L.A. students' grades are rising, but test scores are falling. Why the big disconnect? *Los Angeles Times*. https://www.latimes.com/california/story/2022- 12-22/la-student-reports-card-grades-are-high- test-scores-are-low-why-the-big-disconnect

Feldman, J., & Reeves, D. (2020). Grading during the pandemic: A conversation. *Educational Leadership*, *78*(1), 22–27.

Frey, E. A., & Birnholz, S. (2020). *The rigor gap: Comparing course grades and end–of–course exam results of algebra I and 10th grade English students in Florida* (Working Paper No. 2020–01). The Florida Council of 100.

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., & Staiger, D. O. (2023). The educational consequences of remote and hybrid instruction during the pandemic. *American Economic Review: Insights*, *5*(3), 377– 392. https://doi.org/10.1257/aeri.20220180

Guskey, T. R. (2009, April). *Bound by tradition: Teachers' views of crucial grading and reporting issues.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Corwin Press.

Hamilton, L., Kaufman, J., & Diliberti, M. (2020). *Teaching and Leading Through a Pandemic: Key Findings from the American Educator Panels Spring 2020 COVID-19 Surveys* [Data Note]. RAND Corporation. https://doi.org/10.7249/RRA168-2

Johnson, S. (2021, February 9). California teachers grapple with grading nearly a year after initial school closures. *EdSource*. https://edsource.org/2021/california-teachers-grapple-with-grading-

nearly-a-year-after-initial-school-closures/648376
Kane, T., & Reardon, S. (2023, May 11). Opinion | Parents Don't Understand How Far Behind Their Kids Are in School. *The New York Times*. https://www.nytimes.com/interactive/2023/05/11/opinion/pandemic-learning-losses-steep-but-not-permanent.html
Klinger, C. R., Klinger, D. A., & Volante, L. (2022, April 18). High school grades matter for post-secondary study, but is pandemic assessment fair? *The Conversation*. http://theconversation.com/high-school-grades-matter-for-post-secondary-study-but-is-pandemic-assessment-fair-177187
Kuhfeld, M., Soland, J., & Lewis, K. (2022, June). Test Scores Patterns Across Three COVID-19-Impacted School Years. *American Educational Research Association, 51(7)*. https://doi.org/10.3102/0013189X221109178
Lekholm, A. K., & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation*, *15*(1), 1–23. https://doi.org/10.1080/13803610802470425
Lemov, D. (2023). Your neighborhood school is a national security risk. *Education Next 23(4). https://www.educationnext.org/your-neighborhood-school-national-security-risk-student-achievement-merit-losing-prospects-era-everybody-wins/*
Leonhardt, David. "Where are the students?" *The New York Times*, September 5, 2023, sec. Briefing. https://www.nytimes.com/2023/09/05/briefing/covid-school-absence.html.
Malkus, N., Christensen, C., & Schurz, J. (2020). *School District Responses to the COVID-19 Pandemic: Round 6, Ending the Year of School Closures*. American Enterprise Institute.
Mathews, J. (2022, July 27). Teachers forced to do even more under policies opposed to bad grades. *Washington Post*. https://www.washingtonpost.com/education/2022/07/24/schools-homework-grading-policy-teachers/
Miller, M. W., Rockholt, C., & Jeffries-Simmons, T. (2020). *Student learning and grading guidance*. Office of Superintendent of Public Instruction.
Mozenter, Z. D. (2019). *Essays on the effects of teacher grading standards and other teaching practices* [Ph.D., The University of North Carolina at Chapel Hill]. https://www.proquest.com/docview/2240074332/abstract/246BCFDF8C5D4CD6PQ/1
Opalka, A., & Talkington, B. (2020, May 12). Avoiding Long-Term Harm from "Do No Harm" Grading Policies. *Center on Reinventing Public Education*. https://crpe.org/avoiding-long-term-harm-from-do-no-harm-grading-policies/
Papageorge, N. W., Gershenson, S., & Kang, K. M. (2020). Teacher expectations matter. *The Review of Economics and Statistics*, *102*(2), 234–251. https://doi.org/10.1162/rest_a_00838
Pasco School District Board of Directors. (2020, July 28). *Board of directors' regular meeting*. Pasco School District.
Penuel, W. R. (2021). Reimagining American education: Possible futures for equitable educational assessment. *Phi Delta Kappan*, *103*(4), 54–57. https://doi.org/10.1177/00317217211065829
Reeves, D. (2011). *Elements of grading: A guide to effective practice* (First edition). Solution Tree.
Reeves, D. B. (2004). The case against the zero. *Phi Delta Kappan*, *86*(4), 324–325. https://doi.org/10.1177/003172170408600418
Reich, J., Buttimer, C. J., Fang, A., Hillaire, G., Hirsch, K., Larke, L. R., Littenberg-Tobias, J., Moussapour, R. M., Napier, A., Thompson, M., & Slama, R. (2020). *Remote Learning Guidance From State Education Agencies During the COVID-19 Pandemic: A First Look* [Preprint]. EdArXiv. https://doi.org/10.35542/osf.io/437e2
Reich, J., & Mehta, J. (2020). *Imagining September: Principles and Design Elements for Ambitious Schools During COVID-19* [Preprint]. EdArXiv. https://doi.org/10.35542/osf.io/gqa2w
Reykdal, C. (2020a). *Bulletin No. 024-20*. Office of Superintendent of Public Instruction.
Reykdal, C. (2020b). *Bulletin No. 088-20*. Office of Superintendent of Public Instruction.
Sanchez, E. I. (2023). *Evidence of grade inflation since 2010 in high school English, mathematics, social studies, and science courses*. ACT.
Sanchez, E. I., & Moore, R. (2022). *Grade inflation continues to grow in the past decade*. ACT.
Seattle Public Schools Board of Directors. (2020, August 12). *Board special meeting*. Seattle Public Schools.
Simonetti, I. (2020, June 12). It may be that I never go back to the grading system. *Vox*. https://www.vox.com/first-person/2020/6/12/21288116/coronavirus-covid-19-schools-teachers-grading
Snipes, J., & Finkelstein, N. (2015). *Opening a gateway to college access: Algebra at the right time*. Regional Educational Laboratory West.

Student Learning & Grading Guidance: Frequently Asked Questions. (2020). Office of Superintendent of Public Instruction.

Tan, B. J. (2023). The consequences of letter grades for labor market outcomes and student behavior. *Journal of Labor Economics*, *41*(3), 565–588. https://doi.org/10.1086/719994

Townsley, M. (2020). Grading Principles in Pandemic-Era Learning: Recommendations and Implications for Secondary School Leaders. *Journal of School Administration Research and Development*, *5*(S1), 8–14. https://doi.org/10.32674/jsard.v5iS1.2760

Tyner, A., & Gershenson, S. (2020). Conceptualizing grade inflation. *Economics of Education Review*, *78*, 102037. https://doi.org/10.1016/j.econedurev.2020.102037

Vahle, C., de Araujo, Z., Han, J., & Otten, S. (2023). Teachers' instructional responses to the COVID-19 pandemic. *Teaching and Teacher Education*, 104040. https://doi.org/10.1016/j.tate.2023.104040

Vasquez, A. (2021, December 13). When 'making the grade' takes on new meaning. *EdSource*. https://edsource.org/2021/when-making-the-grade-takes-on-new-meaning/664591

Vázquez Toness. (2023, March 6). Many kids are struggling in school. Do their parents know? *AP NEWS*. https://apnews.com/article/school-report-card-parent-teacher-conference-53b69e145f2dca3a00b50c80004c4a1a

Walker, T. (2021, May). 'Amazing and very challenging': More educators rethink grading. *NEA Today*. https://www.nea.org/advocating-for-change/new-from-nea/amazing-and-very-challenging-more-educators-rethink-grading

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and Test Scores: Accounting for Observed Differences. *Journal of Educational Measurement*, *39*(1), 1–37. https://doi.org/10.1111/j.1745-3984.2002.tb01133.x

Woodruff, D. J., & Ziomek, R. L. (2004). *Differential grading standards among high schools* (ACT Research Report Series No. 2004–2). ACT.

Notes

[1]In 1983, for instance, the authors of "A Nation at Risk" observed that "grades have risen as average student achievement has been declining." (p. 19-20).
[2]By grading standards, we mean the professional judgements teachers use in assigning grades to students and the degree to which the grades align with objective measures of student knowledge. These may be influenced by grading practices (e.g., giving full credit to students who turn in work late, assigning missing work partial credit, or not assigning an F grade to students who turn in all assignments), but could differ because of individual teacher judgments about what quality of student work or mastery of content standards merits a given grade (Betts & Grogger, 2003; Figlio & Lucas, 2004). Given the role grades play in conveying expectations (Guskey & Bailey, 2001; Guskey, 2009) and the link between teacher expectations and student achievement (e.g., Bowden et al., 2023; Papageorge et al., 2020), the relationship between grading standards and achievement is an important question for policy and research.
[3]We use the term loosening to mean that there is a weaker relationship between grades and objective measures of student achievement.
[4]There is relatively little quantitative evidence on the mechanisms for how grading standards might affect achievement, but research at the college level finds a connection between higher anticipated grades and less time spent studying (Babcock, 2010). Other research finds that students on the margin between two letter grades who receive the lower of the two grades (for instance, a student who has an 89% and receives a B) are more likely to take easier courses and receive worse grades in subsequent semesters (Tan, 2023)
[5]Guskey and Link (2019) identify noncognitive elements such as work habits and student progress as being as much of 10% of student grades—the equivalent of a difference between an A and a B.
[6]Grading policy changes brought up conflicts in motivation. In one study, an administrator reportedly told the teacher, "Make them [the students] feel really successful and confident and keep the positivity going" (Vahle et al., 2023, p. 26), but in the same study, teachers reported that after the district implemented a policy that students' grades could not decrease, student participation in online synchronous sessions dropped steeply. The researchers and informants attributed this drop to the loss of grades as a motivational tool.
[7]Of particular interest are places where the meaning of grades changes over time—what Tyner and Gershenson (2020) term "dynamic grade inflation" and when groups of students are held to different standards, or "differential grade inflation." In this paper we focus on the former, but the latter is the focus of our ongoing work.
[8]OSPI specified that, "An 'Incomplete' communicates that a teacher was not able to determine proficiency of the learning standards for the course, which could be attributed to a variety of reasons," and that to assign an incomplete "a teacher must be able to identify the specific standard the student was unable to meet and the steps to demonstrate meeting the standard" (*Student Learning & Grading Guidance*, 2020, p. 3). Students needed to resolve any incomplete grades before graduating high school, and when students with incompletes transferred districts, the sending district was required to "communicate what the student needs to achieve a grade and for the student to successfully resolve the 'Incomplete'" (*Student Learning & Grading Guidance*, 2020, p. 4).
[9]Because of inconsistencies in how courses are identified, we are not able to perfectly distinguish between different levels of algebra classes. We use state course codes and course titles to make reasonable guesses to distinguish between students taking what is generally referred to as "Algebra 1" and students who are taking what is generally referred to as "Algebra 2." We remove all students we identify as taking Algebra 2 because they are likely to be systematically different from their peers taking Algebra 1.
[10]The conversion is as follows: A = 4.0, A- = 3.7, B+ = 3.3, B = 3.0, B- = 2.7; C+ = 2.3, C = 2.0, C- = 1.7, D+ = 1.3, D = 1.0, F = 0.0

[11]The state piloted the Smarter Balanced Assessment (SBA) in 2013-14. In 2014-15, the 11th grade SBA became a high school graduation requirement for some cohorts of high school students; however, 2015-16 was the first year the test was administered to the majority of Algebra 1, Geometry, and Algebra 2 students in high school. Thus, the sample size for the math SBA in 2014-15 is significantly smaller than in subsequent year. This issue impacted the math test, but not the ELA test, which is apparent from the sample sizes in ELA as compared to math.
[12]We focus on Algebra 1, Geometry, and Algebra 2 because those are the math classes most often taken in the tested grades we observe.
[13]Washington did administer a shortened version of the Smarter Balanced Assessment (SBA) in the fall of 2021, but we do not have access to those data. Officials from OSPI noted that this was an unusual test administration: a shorter test at a different time of year (fall rather than spring) taken by fewer students (90% of students instead of 97% of students) (Bazzaz, 2022).
[14]There are a variety of potential reasons for this: students, for instance, may have been absent on testing days, opted out of testing, or have special education status that exempted them from a regular test.
[15]There are some notable shifts in which grade students are taking a subject, such as the increase in students taking Algebra 1 in the 7th and 8th grade that began in the 2019-20 school year. We are uncertain what to make of this shift, but understanding it is part of our ongoing study.
[16]Because districts differ in the number of grades students receive (e.g., on school may give students one grade per trimester while another may give students a single grade for the entire year), we averaged the converted numerical then converted it to the nearest letter grade for all years except for 2019-2020 when we converted to average grades in fall and spring. The conversion is as follows: A: average grade > 3.3; B: average grade ≤ 3.3 and > 2.3; C: average grade ≤ 2.3 and > 1.3; D: average grade ≤ 1.3 and ≥ 1; F: average grade < 1. (3)
[17]Grades in English also increased in the fall of 2019-2020, but not to the same degree as in math.
[18]In some specifications we also include student demographics and/or school district fixed effects, but the inclusion of these controls has little impact on the relationship between grades and test scores.
[19]For instance, students who take Algebra 1 in middle school tend to be higher achieving, but we might worry that the proportion of students taking Algebra 1 in middle schools varies from year to year based on changes in district policies or practices (Clotfelter et al., 2015). [20]Note that in doing these predictions, we are holding constant the grade level in which a student took a course.

We would like to thank OSPI for collecting and providing the administrative data we use. Additional thanks to Mike DeArmond and Emily Morton for comments that improved the paper and to Roddy Theobald and Stephanie Liddle for research assistance.