Received: 11 April 2023 | Accepted: 29 November 2023

DOI: 10.1111/coep.12637

ORIGINAL ARTICLE

CONTEMPORARY ECONOMIC POLICY

Making the grade: The effect of teacher grading standards on student outcomes

Seth Gershenson[1] | **Stephen B. Holt**[2] | **Adam Tyner**[3]

[1]Department of Public Administration and Policy, School of Public Affairs, American University, Washington, District of Columbia, USA

[2]Department of Public Administration and Policy, Rockefeller College, University at Albany - SUNY, Albany, New York, USA

[3]Thomas B. Fordham Institute, Washington, District of Columbia, USA

Correspondence
Seth Gershenson, School of Public Affairs, American University, Washington, DC 20016, USA.
Email: gershens@american.edu

Abstract

One mechanism by which teachers might affect student learning is through the grading standards they set for their classrooms. However, the effects of grading standards on student outcomes are understudied. Using administrative data that links individual students and teachers in Algebra I classrooms from 2006 to 2016, we examine the effects of teachers' grading standards on student learning and attendance. High teacher grading standards increase both contemporaneous student achievement in Algebra I and performance in subsequent math classes. Heterogeneity analyses find that these impacts are positive and similar in size for students of different backgrounds, aptitudes, and school contexts.

KEYWORDS
grade inflation, student achievement, teachers

JEL CLASSIFICATION
I2

1 | INTRODUCTION

A robust consensus in the economics of education agrees that teachers play a pivotal role in shaping students' educational and life trajectories. Effective teachers improve student learning (Chetty et al., 2014a; Rivkin et al., 2005), attendance (Gershenson, 2016; Liu & Loeb, 2021), non-cognitive skills (Jackson, 2018; Kraft, 2019), and long-run outcomes such as college entry and earnings (Chetty et al., 2014b; Gershenson et al., 2022). However, while there is agreement that access to effective teachers is important, relatively little is known about the pedagogical practices that make teachers more or less effective. This gap in the literature limits our ability to train and identify effective teachers.

Various stakeholders have speculated that high grading standards improve student outcomes. This is an intriguing idea, as teachers' expectations and grading practices are likely malleable (De Boer et al., 2018; Pollio & Hochbein, 2015; Quinn, 2020) and teachers can use grades to convey (un)satisfactory progress (or effort) to students and parents (Brookhart et al., 2016; Mechtenberg, 2009). However, economic theory yields mixed predictions regarding whether, and which, students would benefit from high grading standards, as some students may disengage from school rather than increase their efforts (Becker & Rosen, 1992).[1] Whether and for whom high grading standards are beneficial is therefore an empirical question with scant compelling evidence; the current study addresses these questions.

Abbreviations: ELA, English language arts; EOC, end of course; FE, fixed effects; NCERDC, North Carolina Education Research Data Center; OLS, ordinary least squares; SD, standard deviation; VAM, value added model.

Figlio and Lucas (2004) provide the best evidence to date: analyzing 4 years of administrative data from Alachua county schools in Florida, they find that, among third to fifth graders, teachers with high grading standards improved math and reading performance relative to their peers with lower standards and—importantly—that no one is harmed by exposure to high grading standards. The authors identify teachers' grading standards by taking the average of their *B* students' end-of-course (EOC) standardized test scores. The idea is that when a teachers' *B* students had higher standardized test scores than the *B* students of other teachers, they learned more and thus experienced more rigorous grading standards.

The current study formalizes and extends Gershenson (2020), who replicated the basic findings of Figlio and Lucas (2004) in a different context: eighth and ninth grade Algebra I classrooms in North Carolina public schools. Our findings are consistent with those of Figlio and Lucas (2004) in that students typically benefit from exposure to high grading standards and that no students are harmed. Specifically, we extend this literature in three ways. First, we examine effects of grading standards on achievement in other subjects and on student attendance and suspensions, which measure school engagement. Second, we investigate how teachers' grading standards vary across subjects and consider alternative approaches to characterizing a teacher's grading standards. Finally, we investigate how grading standards' effects might vary across students of different *relative* ability levels within the classroom.

2 | DATA AND METHODS

2.1 | Data

We examine the effects of teacher grading standards on student outcomes using administrative data from North Carolina public schools. Specifically, we use student-level data from the North Carolina Education Research Data Center (NCERDC) from 2006 to 2016.[2] The NCERDC data contain transcripts for each student that allow us to link course grades to teacher–student pairings and students' scores on standardized state EOC tests in math in grades 3 through 8 and in Algebra I, Geometry, and Algebra II.[3] We focus on students who took Algebra I in eighth or ninth grade between 2006 and 2016, as these are the grades in which most students take the course. After removing students with incomplete test-score or transcript data, the main analytic sample contains 365,004 unique students and 4455 unique Algebra I teachers spread across 1090 schools and 26,819 Algebra I classrooms.[4]

We focus on the effects of teacher grading standards on three outcomes: contemporaneous performance in Algebra I, performance in subsequent math courses (i.e., Geometry and Algebra II), and student attendance. Algebra I, Geometry, and Algebra II achievement are measured by performance on the standardized EOC test in the subject, which we standardize within grade-year to have mean 0 and standard deviation (SD) 1. In North Carolina, all students in these courses take the EOC test developed by North Carolina's Department of Public Instruction in consultation with input from select teachers and aligned with state curricula standards.[5] Our data include all students not absent or otherwise excused from the tests.

While not an outcome of interest themselves, the course grades assigned by teachers are crucial to the analysis, as they are used in the construction of teacher grading standards. Course grades appear in the transcript data in both numeric and letter form. Numeric grades appear on a scale from 50 to 100. Letter grades appear as A, B, C, D, or F with occasional + or − modifiers.[6] We convert letter grades to numeric values using a simple midpoint rule, such that a B is an 85, a B+ is an 88, and so on. We also observe information on teacher demographics (race and gender), years of experience, licensure (regular math instruction licensing or lateral/provisional licensing), and educational attainment. Finally, we observe administrative information on student demographics (race and gender), economic disadvantage, and prior achievement in math and English language arts (ELA).[7]

2.2 | Measuring teacher grading standards

We follow Figlio and Lucas (2004) in measuring teacher grading standards (*S*) by comparing the more subjective course grades awarded by teachers to the more objective EOC test scores of their students. Intuitively, if one student scores higher on the EOC test than another, but both receive the same grade from their teacher, the student with the higher test score experienced higher grading standards in the sense that they achieved greater content mastery while receiving the same course grade. There are a few ways to formally estimate *S* in practice and both Figlio and Lucas (2004) and

Gershenson (2020) find that the estimated effect of grading standards on student achievement is robust to how this is done. We focus on two main ways, each with their own strengths and weaknesses. In both cases, we exclude current students' EOC scores and grades from the computation to yield a year-specific estimate $\widehat{S}_{jt}$ for teacher j in year t, which avoids the endogeneity concern of current students affecting their teacher's grading standards.[8]

The first approach is to estimate S as the average EOC score of each teacher's B students.[9] This is a simple and transparent approach to defining grading standards. Moreover, it is appealing because nearly all classrooms (and teachers) will have some B students, while certain tracked classrooms may not have many A or C/D students. Here, a higher $\widehat{S}$ (i.e., higher average EOC for B students) means higher grading standards.

The second and preferred approach is a regression-based, value-added measure where S_{jt} is operationalized as a set of teacher-year indicators (fixed effects) in a regression of EOC scores on S_{jt} and course grades, again omitting year-t students from the regression:

$$EOC_{ij,-t} = S_{jt} + f(Grade)_{ij,-t} + \varepsilon_{ij,-t}, \forall t, \tag{1}$$

where i, j, and t index students, teachers, and years, respectively. The S in Equation (1) identifies the residualized teacher-year specific average EOC score of all non-t students assigned to teacher j after accounting for the grade given by teacher j to student i. We estimate Equation (1) both by OLS and using a shrinkage estimator that adjusts for variation across teachers in the number of students taught (Chetty et al., 2014a; Stepner, 2013). Once again, higher values of $\widehat{S}_{jt}$ indicate higher grading standards, as students of teacher j demonstrate more mastery of the material on EOC tests than do students of other teachers, despite receiving the same course grade. This regression-based approach is attractive because it uses data on all students (not just B students) and because the estimates can be shrunken to account for teacher-specific variance in the precision of value-added estimation.

2.3 | Identifying the impact of teacher grading standards

We wish to estimate the causal relationship between students' exposure to high teacher grading standards and their short- and medium-term learning outcomes. Again following Figlio and Lucas (2004), we identify this relationship by including $\widehat{S}$ as a teacher characteristic in a standard value-added model of the education production function (Harris & Sass, 2011; Pope, 2019). Specifically, we estimate:

$$Y_{ijgst} = \beta\widehat{S}_{jt} + \delta X_{it} + \varphi M_{jt} + \theta_{gst} + \varepsilon_{ijgst}, \tag{2}$$

where i, j, and t again index students, teachers, and years while g and s index grade and school, respectively. In Equation (2), X is a vector of controls for student characteristics (race, gender, prior achievement); M is a vector of teacher controls (race, gender, educational attainment, years of experience, license held, and math value added); and θ represents a grade-by-school-by-year fixed effect that ensures we compare the outcomes of students in the same school and same grade at the same time but who are exposed to different grading standards. Math value added is computed using an equation and estimation procedure similar to that described in Equation (1) (Chetty et al., 2014a; Stepner, 2013); it is standardized to have mean zero and SD one. The inclusion of students' prior math achievement in X accounts for variation in pre-existing student ability and potential sorting of students to particular classrooms (Chetty et al., 2014a). Because we focus on students taking Algebra I in eighth or ninth grade, we use EOC math exam scores from the prior grade (seventh or eighth grade, respectively).[10]

The parameter of interest is β, which represents the effect of a one unit increase in teacher grading standards (measured in EOC SD). In practice, we replace $\widehat{S}$ with a set of indicators for which quartile of the distribution of grading standards a teacher falls in, though the main results are robust to using the continuous measure. This makes our estimates of β more readily interpretable and allows for nonlinear effects of teacher grading standards.

Finally, while Equation (2) is in many ways a standard value-added model now common in economic research, two nuances merit further discussion. First, the predictor of interest, $\widehat{S}$, is an estimate. The usual OLS standard errors of β will therefore be too small. Accordingly, we compute standard errors using a bootstrap procedure (1000 replications). We allow for serial correlation within schools by implementing a block bootstrap at the school-grade-year level.[11]

Second, the school-by-grade-by-year fixed effect (θ_{gst}) plays an important role in our analysis, as it ensures that we are comparing students who have different Algebra I teachers with different levels of $\widehat{S}$ but are taking Algebra I in the

same school year, at the same school, and in the same grade. Comparing students studying the same material, in the same environment, and at the same academic stage ensures that the only variation contributing to β_n comes from exposure to different classrooms and avoids confounders that might jointly affect teacher assignment and EOC scores. A necessary trade-off of this research design is that because it relies on comparisons of teachers in the same school teaching Algebra I for the same grade at the same time, the estimate is identified by students in schools with multiple Algebra I teachers in a given year.[12] Accordingly, we verify that the main findings are robust to replacing θ_{gst} with separate school, grade, and year FE.

2.4 | Summary statistics

Table 1 summarizes the Algebra I teachers and students who comprise the analytic sample. Summary statistics are reported both overall and separately by quartile of the grading standards distribution. Columns 1–3 summarize the teachers, who are about 80% white and 30% male. Teachers in the top quartile (toughest standards) awarded higher

TABLE 1 North Carolina Algebra 1 teacher-years (eighth or ninth grade), 2006–2016.

	Teachers			Students		
	All (1)	Bottom Q (2)	Top Q (3)	All (4)	Bottom Q (5)	Top Q (6)
Test score	−0.06 (0.67)	−0.50 (0.62)	0.47*** (0.57)	0.06 (0.97)	−0.40 (0.93)	0.55*** (0.90)
Course grade	79.07 (9.45)	76.62 (9.58)	82.16*** (8.76)	80.12 (12.63)	77.93 (12.82)	82.94*** (11.76)
Lagged math score	0.14 (0.76)	−0.18 (0.73)	0.55*** (0.74)	0.23 (0.93)	−0.07 (0.90)	0.62*** (0.89)
Student absences	6.59 (4.72)	7.86 (6.10)	5.54*** (3.91)	5.98 (6.90)	6.78 (8.26)	5.17*** (5.55)
Any suspensions	0.16	0.19	0.17	0.13	0.15	0.12***
White	0.81	0.68	0.93***	0.50	0.37	0.62***
Black	0.15	0.28	0.05***	0.22	0.33	0.12***
Hispanic	0.01	0.01	0.01	0.10	0.12	0.08***
Asian	0.01	0.02	0.01***	0.02	0.02	0.03***
Native American	0.01	0.01	0.01	0.01	0.01	0.01***
Other race	0.01	0.01	0.00***	0.04	0.04	0.03***
Male	0.29	0.33	0.24***	0.49	0.50	0.49***
Lateral license	0.09	0.13	0.06***	–	–	–
Provisional license	0.02	0.04	0.01***	–	–	–
Teacher experience	9.50 (5.65)	8.47 (5.56)	10.40*** (5.64)	– –	– –	– –
Advanced degree	0.28	0.24	0.35***	–	–	–
Free or reduced-price lunch	–	–	–	0.44	0.56	0.31***
Observations	15,854	4708	3497	471,997	118,185	117,713

Note: Columns 1–3 observations are teacher-years and columns 4–6 observations are student-years. Standard deviations in parentheses. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$ for t-test of difference in means between column 3 and column 2 and column 6 and column 5, respectively. Q refers to quartile of VAM-estimated teacher grading standards. Test score refers to the average within teacher-year student score on the North Carolina Algebra I end of course tests (standardized by grade-year). Course grade in columns 1–3 refers to the within teacher-year average numeric grade assigned to students in Algebra I and in columns 5 and 6 refers to the average numeric grade assigned to students in the analytic sample. The lagged math score in columns 1–3 refers to the within teacher-year average math achievement in the prior year of students assigned to teacher j in year t (standardized by grade-year).

course grades and had students with fewer absences and higher test scores, on average, than their counterparts with lower grading standards. Teachers with higher grading standards were disproportionately white, female, and more experienced. The latter difference suggests that grading standards may be malleable and also highlights the importance of controlling for other teacher attributes in Equation (2) in order to isolate the effect of grading standards. Moreover, mounting evidence that teacher effectiveness grows with experience (Papay & Kraft, 2015; Wiswall, 2013) suggests that changing grading standards might be one mechanism through which experience translates into effectiveness.

Columns 4–6 of Table 1 summarize students. About half the Algebra I students in the sample are white and the average student received a *B*− grade. White, higher performing, and more advantaged students were more likely to have a high-standards Algebra I teacher. Differences in the lagged test scores of students assigned to more and less lenient graders further highlight the nonrandom sorting of students and teachers and accentuate the need for the value-added model characterized by Equation (2).

Finally, the statistical properties of regression-based measures of teachers' grading standards are less established than other measures of teacher effects, such as value-added measures. To give a sense of how stable teacher grading standards are over time, Table 2 reports intertemporal Spearman rank correlation coefficients for each 2-year pair in our analytic sample. We also report the 2-year correlation coefficients of teacher value-added estimates as a point of comparison,

TABLE 2 Stability of Algebra I teacher grading standards and value-added, 2007–2016.

Two-year teacher effects	Grading standards	VAM	*N*
2007–2008	0.73***	0.45***	1012
2008–2009	0.67***	0.43***	726
2009–2010	0.72***	0.50***	845
2010–2011	0.64***	0.54***	1318
2011–2012	0.71***	0.59***	1209
2012–2013	0.63***	0.44***	735
2013–2014	0.68***	0.42***	657
2014–2015	0.67***	0.49***	867
2015–2016	0.70***	0.59***	1108

Note: Spearman rank correlations are reported. Two-year teacher effects are estimated using teacher-year-specific rankings of teacher effects from each year in the two consecutive year pairs detailed in the rows.

***$p < .01$.

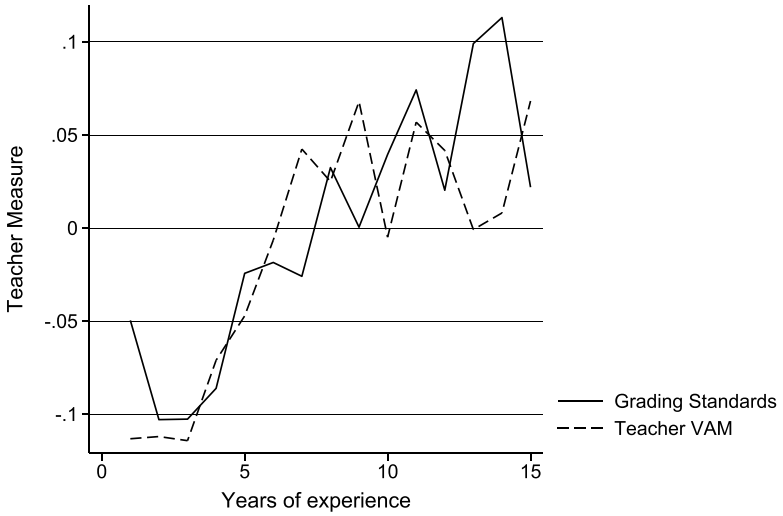

FIGURE 1 Stability of teacher grading standards and teacher effectiveness. Plot shows the average value of *S* (solid line) and the average value-added score in math (dashed line) by years of experience.

which are similar to what is found in the broader literature (Loeb & Candelaria, 2012). Our preferred measure of grading standards is actually more stable year to year than value-added measures of teacher effectiveness. Moreover, Figure 1 plots the average teacher value-added and average grading standards estimate by teachers' years of experience. Teacher grading standards follow a similar pattern as the value-added measures: they both increase with experience, which is consistent with the idea that grading standards measure a fairly stable, teacher-specific trait that evolves over time.

3 | RESULTS

3.1 | Main results

Table 3 presents estimates of several versions of the baseline specification given in Equation (2), where the outcome is contemporaneous performance on the Algebra I EOC exam. The models in Table 4 condition on teachers' math value-added scores, but are otherwise identical to those in Table 3. In both tables, and in all model specifications, the estimated effect of rigorous grading standards is positive and strongly statistically significant. The estimated effect of a one standard deviation increase in teacher effectiveness is about 9% of a test-score SD, which is consistent with the existing value-added literature (Hanushek & Rivkin, 2010).

Comparing the estimates in Tables 3 and 4 we see that adjusting for teachers' effectiveness (as measured by math value-added) reduces the estimated effect of grading standards by about 50%, though the effect of grading standards remains economically and statistically significant. This suggests that grading standards are a dimension of teacher effectiveness that is partly, but not entirely, captured by traditional value-added measures. Moreover, it suggests that the documented correlation between grading standards and student outcomes is not the spurious result of more effective teachers also employing more rigorous grading standards.[13]

Columns 1 and 2 estimate the baseline model, using the preferred VAM-based shrinkage estimates of grading standards. Here, teachers are placed into quartiles of the grading standards distribution. Column 1 omits the student

TABLE 3 Effect of teacher grading standards on Algebra I test scores.

	VAM W/Shrinkage				Avg. "B"	VAM 2
	(1)	(2)	(3)	(4)	(5)	(6)
Bottom Q	(Omitted)					
Lower-middle Q	0.06***	0.06***		0.07***	0.08***	0.07***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Upper-middle Q	0.11***	0.10***		0.11***	0.14***	0.13***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Top Q	0.19***	0.18***		0.19***	0.20***	0.17***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Teacher standards score			0.25***			
			(0.01)			
Adjusted R^2	0.45	0.46	0.46	0.45	0.46	0.46
Observations	471,997	471,997	471,997	471,997	471,990	471,997
Student X	No	Yes	Yes	Yes	Yes	Yes
Teacher X	No	Yes	Yes	Yes	Yes	Yes
School-grade-year FE	Yes	Yes	Yes	No	Yes	Yes
School FE	No	No	No	Yes	No	No

Note: Bootstrapped standard errors (1000 replications) clustered at the school-grade-year level (columns 1–3, 5, and 6) and school-level (column 4) in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Q refers to quartile of teacher grading standards student i faces in Algebra 1. Columns 1 through 4 measure grading standards using a regression-based value-added model with shrinkage to account for classroom size variation across teacher-years. Average "B" in column 5 measures teacher grading standards as the leave-year out average score of students receiving a "B" grade from teacher j. VAM 2 in column 6 measures grading standards using a regression-based value-added model without shrinkage adjustment.

TABLE 4 Effect of teacher grading standards on Algebra I test scores controlling for value-added.

	VAM W/Shrinkage				Avg. "B"	VAM 2
	(1)	(2)	(3)	(4)	(5)	(6)
Bottom Q	(Omitted)					
Lower-middle Q	0.03***	0.03***		0.04***	0.05***	0.04***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Upper-middle Q	0.06***	0.05***		0.07***	0.08***	0.07***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Top Q	0.10***	0.09***		0.12***	0.09***	0.08***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Teacher standards score			0.11***			
			(0.01)			
Teacher VAM			0.08***			
			(0.00)			
Bottom Q VAM	(Omitted)					
Lower-middle Q VAM	0.07***	0.06***		0.06***	0.06***	0.06***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Upper-middle Q VAM	0.12***	0.12***		0.10***	0.11***	0.11***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Top Q VAM	0.20***	0.20***		0.18***	0.18***	0.20***
	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Adjusted R^2	0.45	0.46	0.46	0.45	0.46	0.46
Observations	462,163	462,163	462,163	462,163	462,156	462,163
Student X	No	Yes	Yes	Yes	Yes	Yes
Teacher X	No	Yes	Yes	Yes	Yes	Yes
School-grade-year FE	Yes	Yes	Yes	No	Yes	Yes
School FE	No	No	No	Yes	No	No

Note: Bootstrapped standard errors (1000 replications) clustered at the school-grade-year level (columns 1–3, 5, and 6) and school-level (column 4) in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Q refers to quartile of teacher grading standards student i faces in Algebra 1. Columns 1 through 4 measure grading standards using a regression-based value-added model with shrinkage to account for classroom size variation across teacher-years. Average "B" in column 5 measures teacher grading standards as the leave-year out average score of students receiving a "B" grade from teacher j. VAM 2 in column 6 measures grading standards using a regression-based value-added model without shrinkage adjustment. Teacher value-added in math included as a control (VAM; also estimated using a regression-based model with shrinkage).

and teacher controls and shows a clear positive gradient in the relationship between grading standards and student achievement: the estimated impact of grading standards is positive, statistically significant, and approximately linear. Column 2 adds the student and teacher controls to the model, yielding nearly identical estimates. This is true in both tables and, again, suggests that there is an effect of grading standards on student achievement that is separate from other observable dimensions of teacher quality (e.g., experience). It also suggests that the lagged test score does a good job of controlling for nonrandom sorting into classrooms.

Column 3 replaces the grading standard quartile indicators with a single, continuous measure of $\widehat{S}$. The estimate here is once again positive and strongly statistically significant. Given the standard deviation of our regression-based measure of $\widehat{S}$ (0.37), the results in Panel A of column 3 suggest that a student taught by a teacher one-standard deviation above the mean in terms of grading standards scores 0.09 standard deviations higher on end of course exams in Algebra I (0.37×0.25); Table 4 suggests a 0.04 SD increase, holding teacher math value-added constant. Column 4 returns to the preferred categorical treatment specification but replaces the school-by-grade-by-year FE with separate FE for school, grade, and year. This allows smaller schools to contribute identifying variation to the estimates, and once again the

TABLE 5 Effect of teacher grading standards on other outcomes.

	Geometry	Algebra II	Absences	Excused	Unexcused	Suspended	Spillover	
							Geometry	Algebra II
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bottom Q	(Omitted)							
Lower-middle Q	−0.00	0.02	−0.00	−0.01	−0.04*	−0.01**	0.03	0.16***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.08)	(0.05)
Upper-middle Q	0.03*	0.08***	−0.01	−0.03	−0.05*	−0.02*	0.19***	0.10*
	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.07)	(0.06)
Top Q	0.04**	0.07***	−0.01	−0.05*	−0.06**	−0.02*	0.24***	0.24***
	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.08)	(0.07)
All controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.34	0.25				0.09	0.32	0.23
Pseudo R^2			0.19	0.14	0.19			
Observations	48,521	88,084	416,662	104,896	104,847	65,018	17,035	41,774

Note: Bootstrapped standard errors (1000 replications) clustered at the school-grade-year level (columns 1, 2, 6, and 7) in parentheses; robust standard errors clustered at the school-level (columns 3–5) in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Data for Geometry tests only available for years 2007–2010; Algebra II test data only available from 2007 to 2011. Suspensions data drawn from 2007 to 2009. Grading standards measured using a regression-based value-added model with shrinkage to account for classroom size variation across teacher-year. In columns 1–5, Q refers to quartile of teacher grading standards student *i* faces in Algebra 1. Columns 3 through 5 present Poisson fixed-effect model estimates. Column 6 presents LPM estimates on a binary indicator equal to 1 if a student is ever suspended during the school year (includes control for lagged count of suspensions the prior year). In columns 7 and 8, Q refers to quartile of grading standards of teacher *j* in Algebra I classes and the sample includes only Algebra II and Geometry classes taught by teacher *j* in the same school-year as Algebra I classes taught by teacher *j*.

estimates are quite similar to those of the preferred specification. Finally, columns 5 and 6 of Table 3 use alternative definitions of grading standards: the average for B students and the unshrunken VAM estimates of grading standards, respectively. Once again, all of these different approaches yield similar results: each quartile increase in grading standards improves student achievement by about 5%–8% of a test-score SD (or 2%–4% conditional on math value added).

To place these estimates in context, consider that a meta-analysis of 314 randomized controlled trials of interventions that targeted math skills suggest that the most conservative estimate from Table 3, 16% of a SD in column 6 from moving from the bottom to top quartiles, falls at the 70th percentile of documented effect sizes (Kraft, 2020). A more marginal move from one quartile to another yields an effect of about 5% of a test-score SD, which is near the median of the effect size distribution. Thus these estimates are large enough to be economically meaningful but not so large as to be unbelievable. Indeed, even a modest increase in teacher grading standards, raising standards above the lowest quartile, would be equivalent to replacing an average teacher with a teacher at the 70th percentile of the effectiveness distribution (Herrmann & Rockoff, 2012; Kraft, 2020).

Having seen that grading standards improve current performance, a natural follow-up question is whether those effects persist in subsequent years. We investigate this question by re-estimating the baseline model for the subset of students who took Geometry or Algebra II in 10th or 11th grade (the next two courses in North Carolina's math sequence) using EOC scores in those subjects as the outcome.[14] The treatment of interest remains the Algebra I teacher's grading standards. We also look at effects on contemporaneous attendance and "spillover effects" of teacher *j*'s Algebra I grading standards on other math classes taught contemporaneously by teacher *j*. These estimates are reported in Table 5.

Columns 1 and 2 show significant effects of Algebra I teachers' grading standards on math achievement one and 2 years later in Geometry and Algebra II, respectively. These estimates generally resemble those previously observed in Algebra I classes, as exposure to higher standards leads to significantly higher test scores in subsequent math courses. These are smaller than the contemporaneous effects on Algebra I test scores, which is consistent with existing evidence of "fade out" over time in teacher effects (Jacob et al., 2010). However, two subtle differences emerge. First, the effects are no longer approximately linear, as there are diminishing returns to the highest standards. The biggest increase comes from moving from the bottom of the distribution to the third quartile. Second, the effects are about twice as large in Algebra II as in Geometry, despite Geometry typically coming first in the math sequence. While this contradicts the

aforementioned evidence of fade out, the reason is likely that the skills and content knowledge needed for success in Algebra I align more closely with those in Algebra II than Geometry. All told, this suggests that the contemporaneous gains in Algebra I learning due to high grading standards were the result of some real learning that persisted over time and was not exclusively due to cramming for the EOC exam or changing study habits and engagement in a short-lived way.

Increased effort and engagement could still be channels through which grading standards improve student outcomes (Figlio & Lucas, 2004); similarly, decreased effort and engagement could be unintended consequences of high standards (Betts & Grogger, 2003; Lillard & DeCicca, 2001). In columns 3 through 5 of Table 5 we investigate the potential for high grading standards to impact other academic behaviors, such as effort and engagement in school, using student absenteeism as a proxy that is also interesting in its own right: student attendance is an educational input that affects achievement and longer-term outcomes like graduation and college going (Liu et al., 2021) as well as an output affected by teachers and other inputs such as class size (Gershenson, 2016; Liu & Loeb, 2021; Tran & Gershenson, 2021).

We measure absenteeism as the count of total days absent and as separate counts of excused and unexcused absences. Importantly, we find no evidence that high grading standards increase absenteeism. In fact, consistent with earlier results showing increased academic achievement, exposure to high grading standards seems to slightly reduce absences. These effects are driven by a reduction in unexcused absences, as seen in column 5.

Finally, columns 6 and 7 of Table 5 show the "spillover" effect of teacher grading standards into other classes taught by the same teacher. One potential concern is that teachers' grading standards are course- or classroom-specific and not part of their general approach to teaching. We show that among the subset of teachers who teach both Algebra I and Geometry (column 6) or Algebra II (column 7) contemporaneously, teachers with higher grading standards in Algebra I also have a large, positive effect on student achievement in their other math classes. This suggests that teachers with high grading standards tend to maintain those standards in all of the classes they teach.

3.2 | Distributional effects and heterogeneity

High grading standards may differentially affect some student subgroups. Indeed, one of the main arguments against high grading standards is that low-performing students and those from historically disadvantaged groups may become discouraged and disengaged from school when exposed to high standards and the ensuing lower grades. Murphy and Weinhardt (2020) find that students' ability relative to their peers can significantly affect achievement and confidence. Similarly, relatively strong students may benefit from exposure to high grading standards that keep them engaged and challenged. Consistent with this possibility, Betts and Grogger (2003) found that high-ability students and White students received the largest benefit from high grading standards. Alternatively, Figlio and Lucas (2004) found only modest differences in the effect of grading standards between high- and low-ability students, consistent with the possibility that high grading standards induce more effort among both high achievers to secure a good grade and low achievers to avoid a failing grade. In Norway, Bonesrønning (2004) found small effects on low achieving students but large effects on high achieving students, again consistent with the possibility that high standards induce effort, even among high achievers. We test for heterogeneous effects by estimating Equation (2) separately by race, gender, locale, and relative math ability. These estimates and their confidence intervals are plotted in Figure 2.

In Figure 2, we estimate versions of Equation (2) that include only an indicator for the teacher being in the top quartile of the grading-standards distribution and thus compare the students of these teachers to those of all others.[15] All told, we estimate the model for 20 distinct subsamples and once using the full analytic sample. It is immediately apparent that there is not much heterogeneity: all of the point estimates are between 0.05 and 0.15, most are in the middle of that range at about 0.10, and all are strongly statistically significant. Particularly in terms of demographics, there are no significant differences by race or gender.

There are a few ways to conceptualize relative math ability and how that might moderate the impact of high grading standards. We consider three definitions, which all use the previous year's standardized test scores, and again all yield similar results: within-classroom math EOC ranks, higher math EOC than ELA EOC, and within-state math EOC ranks. For the two rank-based definitions we focus on the top and bottom terciles.[16] These estimates are all qualitatively similar and not statistically distinguishable from one another. The largest practical difference here is between the top and bottom terciles in the statewide ranking, where top-tercile performers benefit about 0.025 SD more from exposure to high grading standards than bottom-tercile students, though again this difference is not statistically significant. The

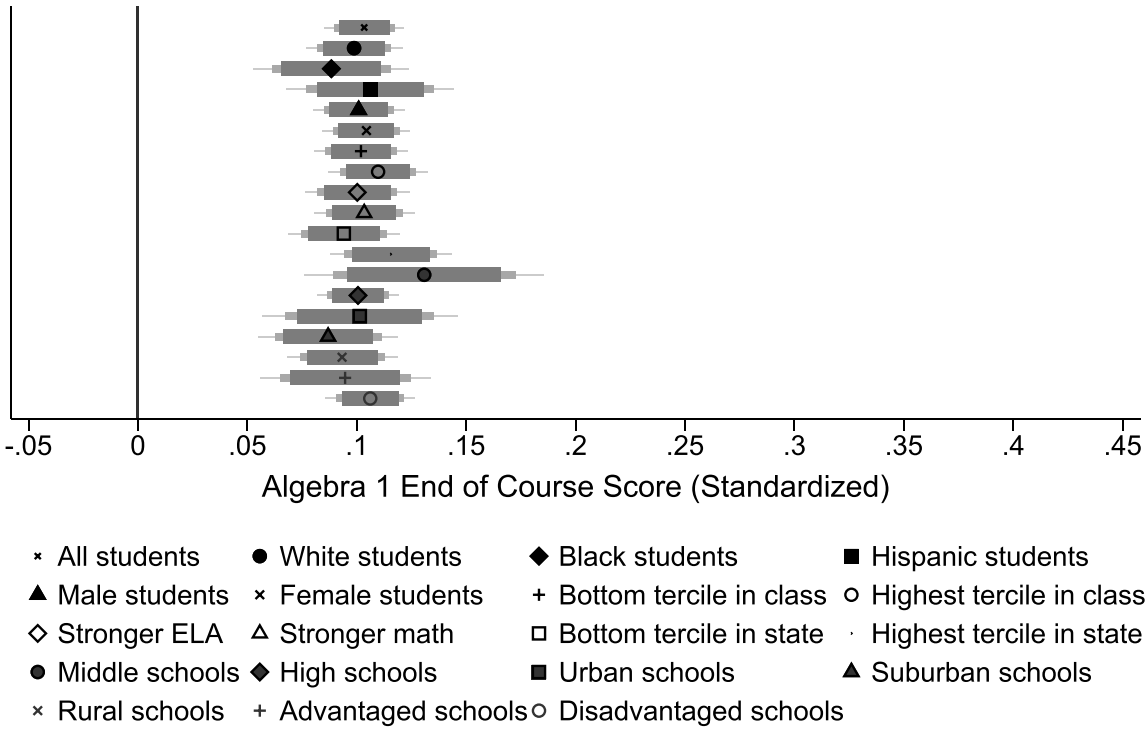

FIGURE 2 Effect of teacher in the top quartile of grading standards across subsamples. Figure plots coefficient of teachers in the top quartile of grading standards, measured using a regression-based value-added model with shrinkage to account for classroom size variation across teacher-year, relative to all other quartiles of standards on achievement in Algebra I courses across subgroups of students and school contexts. Estimated using full model with all controls (column 2 of Table 3). 90%, 95%, and 99% confidence intervals plotted in descending order of line thickness. Bottom and highest terciles in class refers to students in the bottom and top terciles of math achievement in the prior year relative to their peers in the same Algebra I classroom. Stronger ELA (math) refers to students who had higher achievement in ELA (math) relative to math (ELA) in the year prior to taking Algebra I. Bottom and highest terciles in state refers to students in the lowest and highest terciles in terms of math ability statewide by grade-year prior to taking Algebra I. Advantaged schools refers to schools in which >50% of students are eligible for free or reduced price lunch. Disadvantaged schools refers to schools in which >50% of students are eligible for free or reduced price lunch.

largest difference we see is between middle and high schools, with the effect of grading standards being about 0.03 SD larger in the former. This could be because by definition, students taking Algebra I in middle school are more advanced in the sense that they are taking Algebra I early, and thus react better to high standards. This interpretation is also consistent with the results for high performers. Still, that all students in all types of schools benefit, on average, from exposure to high grading standards is consistent with the arguments posited by Figlio and Lucas (2004) and Feltovich et al. (2002) that standards induce effort from both high- and low-ability students within a classroom by increasing the risk of receiving a lower grade. These findings are also consistent with prior empirical evidence (Betts & Grogger, 2003; Figlio & Lucas, 2004).

4 | CONCLUSION

Teachers' expectations for students can exert tremendous influence on their educational achievement and attainment (Papageorge et al., 2020). One way that teachers operationalize high expectations is via high grading standards. We show that in North Carolina, teachers with high grading standards have large, significant, positive effects on students' performance in Algebra I and in subsequent math courses. Smaller, marginally significant reductions in student

absenteeism are observed as well. This is true across all student demographic groups, school types, and achievement levels; no students appear to be harmed by exposure to high standards. The effects on absences and subsequent math classes are consistent with the idea that exposure to high grading standards increases student effort and general engagement with school and the change lasts beyond the current class.

These findings are related to two literatures in the economics of education. First, that high grading standards improve achievement is consistent with evidence from North Carolina that teachers' educational expectations of students improves test scores (Hill & Jones, 2021) and from a nationally representative sample that teacher expectations affect college completion (Papageorge et al., 2020). Similarly, these results align with studies of the long-run effects of exposure to biased teachers, where biases are identified by comparing the grades (scores) assigned to blind and non-blind tests (Lavy & Sand, 2018; Terrier, 2020).

Second, grading standards are fundamentally linked to the phenomenon of grade inflation. Historically, grade inflation (low grading standards) is more prevalent in schools serving economically disadvantaged communities (Tyner & Gershenson, 2020). Coupled with our findings, this suggests that disproportionate exposure to lax grading standards contributes to socio-economic disparities in educational outcomes. More broadly, school leaders and policy makers should be concerned that grade inflation seems to be increasing at all types of K-12 schools as well as in colleges and universities (Denning et al., 2022; Tyner & Gershenson, 2020). Besides reducing achievement and effort and obfuscating the areas in which students could strive for improvement, this reduces the signal-to-noise ratio of grades as a signal in the labor market and might influence students' choices about major, graduate school, and occupation.

One possible response, and a way to introduce or increase standards, is to require that students take certain courses or curricula in high school. This has been done in a handful of states with relatively little net benefit, as any modest achievement gains are often offset by decreases in high school completion rates (Lillard & DeCicca, 2001). Jacob et al. (2017), for example, study the implementation of the Michigan Merit Curriculum and find modest gains in ACT scores coupled with small drops in graduation rates among relatively low performing students. Similarly, many states have included exit exams as a requirement for graduation, which also tend to reduce graduation rates among lower-performing and economically disadvantaged students (Dee & Jacob, 2006; Papay et al., 2022). Because we find that all students benefit from high grading standards, the more moderate policy change of increasing grading standards and relying on continuous rather than binary performance indicators (i.e., course grades rather than a high-stakes pass/fail test) can capture the benefits of high standards for all students, including high-achievers, while avoiding the negative consequences of the latter.

ACKNOWLEDGMENTS

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the North Carolina Educational Research Data Center (NCERDC). Restrictions apply to the availability of this data, which may be requested and obtained from the NCERDC in accordance with their policies. Procedures for requesting and obtaining the data are here: https://childandfamilypolicy.duke.edu/wp-content/uploads/sites/2/2023/04/NCERDC_Procedures-for-Obtaining-Data_23.pdf. The code and do files that prepare the data for, and run, the analyses in this paper are available in Gershenson et al. (2023).

ENDNOTES

[1] On the one hand, high grading standards raise the cost of low effort and low performance in school, which might induce both students and parents to become more engaged. Similarly, low grading standards might make it difficult for students to identify areas for improvement or create a false sense of achievement that leads to complacency. More nefariously, students might interpret low standards as their teacher not believing in their ability to succeed, creating a self-fulfilling prophecy in which students fail to believe in their own potential (Papageorge et al., 2020). On the other hand, high grading standards that yield low grades might discourage students, particularly those from historically marginalized backgrounds, and cause them to further disengage from academics.

[2] NCERDC provides access to administrative data from North Carolina public schools to researchers with a data use agreement. More on the data at NCERDC and the process for accessing the data for research can be found here: https://childandfamilypolicy.duke.edu/north-carolina-education-research-data/. Stata do files that describe the cleaning and analysis of the data can be found at the OpenICPSR repository (Gershenson et al., 2023).

[3] In 2014, North Carolina implemented the Common Core Curriculum and Standards. Under the Common Core Standards, Algebra I aligns approximately with Math I, and Geometry and Algebra II align with Math II and Math III respectively. Unfortunately, NCERDC does not provide end of course test scores for Math II or III, which restricts our Geometry and Algebra II samples to 2006 through 2014.

[4] Our analytic sample differs from (Gershenson, 2020) because we impute years of experience for teachers missing experience information using the number of years we observe them within the North Carolina public school system.

[5] For more information, see https://www.dpi.nc.gov/districts-schools/testing-and-school-accountability/state-tests/end-course-eoc

[6] It is unclear why some schools report letter grades and others report numeric grades. In any case, there are no systematic differences by reporting type in the student socio-demographic composition of schools and separate analyses by reporting type yield similar results (Tyner & Gershenson, 2020).

[7] Economic disadvantage is defined as eligible for the means-tested free or reduced price lunch program.

[8] One implication of excluding current students is that all teachers who only appear in the data for a single year are not included in our sample.

[9] Formally, $\widehat{S}_{jt} = \frac{\Sigma EOC_{i,j,-i}}{N_{j,-t}} | Grade_{i,j,-t} = B$, for all j, t.

[10] North Carolina's Department of Public Instruction requires students to take standardized math and ELA exams at the end of each grade from fourth grade through eighth grade.

[11] The level of clustering does not matter from a practical standpoint: the main results are strongly statistically significant regardless of whether we cluster by classroom, teacher, school year, or school.

[12] Note that θ_{gst} necessarily subsumes the school, grade, and year fixed effects typically included in value-added models. School- and grade-level variables are similarly redundant.

[13] The Spearman correlation coefficient between teacher math value-added and grading standards is 0.517.

[14] Data for Geometry tests are only available for years 2007–2010 and Algebra II test data are only available from 2007 to 2011. Table A4 in the Appendix estimates our main model of Algebra I teacher grading standards on Algebra I achievement using each subsample of non-missing data on alternative outcomes (i.e., Geometry, Algebra II, and attendance) to demonstrate that the main effects are consistent across these changes in sample due to data availability.

[15] See Tables A1–A3 in the appendix for table versions of these estimates.

[16] We use terciles because in a typical class of 20 or fewer students, student groupings become quite small with additional bins.

REFERENCES

Becker, W.E. & Rosen, S. (1992) The learning effect of assessment and evaluation in high school. *Economics of Education Review*, 11(2), 107–118. Available from: https://doi.org/10.1016/0272-7757(92)90002-k

Betts, J.R. & Grogger, J. (2003) The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22(4), 343–352. Available from: https://doi.org/10.1016/s0272-7757(02)00059-6

Bonesrønning, H. (2004) Do the teachers' grading practices affect student achievement? *Education Economics*, 12(2), 151–167. Available from: https://doi.org/10.1080/0964529042000239168

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

Chetty, R., Friedman, J.N. & Rockoff, J.E. (2014a) Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. *The American Economic Review*, 104(9), 2593–2632. Available from: https://doi.org/10.1257/aer.104.9.2593

Chetty, R., Friedman, J.N. & Rockoff, J.E. (2014b) Measuring the impacts of teachers II: teacher value-added and student outcomes in adulthood. *The American Economic Review*, 104(9), 2633–2679. Available from: https://doi.org/10.1257/aer.104.9.2633

De Boer, H., Timmermans, A.C. & Van Der Werf, M.P. (2018) The effects of teacher expectation interventions on teachers' expectations and student achievement: narrative review and meta-analysis. *Educational Research and Evaluation*, 24(3–5), 180–200. Available from: https://doi.org/10.1080/13803611.2018.1550834

Dee, T. & Jacob, B. (2006) Do high school exit exams influence educational attainment or labor market performance? *National Bureau of Economic Research*. Available from: https://doi.org/10.3386/w12199

Denning, J.T., Eide, E.R., Mumford, K.J., Patterson, R.W. & Warnick, M. (2022) Why have college completion rates increased? *American Economic Journal: Applied Economics*, 14(3), 1–29. Available from: https://doi.org/10.1257/app.20200525

Feltovich, N., Harbaugh, R. & To, T. (2002) Too cool for school? Signalling and countersignalling. *The RAND Journal of Economics*, 33(4), 630–649. Available from: https://doi.org/10.2307/3087478

Figlio, D.N. & Lucas, M.E. (2004) Do high grading standards affect student performance? *Journal of Public Economics*, 88(9–10), 1815–1834. Available from: https://doi.org/10.1016/s0047-2727(03)00039-2

Gershenson, S. (2016) Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11(2), 125–149. Available from: https://doi.org/10.1162/edfp_a_00180

Gershenson, S. (2020) *Great expectations: the impact of rigorous grading practices on student achievement*. Washington, DC: Thomas B. Fordham Institute.

Gershenson, S., Hart, C., Hyman, J., Lindsay, C.A. & Papageorge, N.W. (2022) The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy*, 14(4), 300–342. Available from: https://doi.org/10.1257/pol.20190573

Gershenson, S., Holt, S. & Tyner, A. (2023) *COEP replication package for "Making the Grade: The Effect of Teacher Grading Standards on Student Outcomes"*. Ann Arbor: Inter-university Consortium for Political and Social Research. Available from: https://doi.org/10.3886/E195724V1

Hanushek, E.A. & Rivkin, S.G. (2010) Generalizations about using value-added measures of teacher quality. *The American Economic Review*, 100(2), 267–271. Available from: https://doi.org/10.1257/aer.100.2.267

Harris, D.N. & Sass, T.R. (2011) Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7–8), 798–812. Available from: https://doi.org/10.1016/j.jpubeco.2010.11.009

Herrmann, M.A. & Rockoff, J.E. (2012) Worker absence and productivity: evidence from teaching. *Journal of Labor Economics*, 30(4), 749–782. Available from: https://doi.org/10.1086/666537

Hill, A.J. & Jones, D.B. (2021) Self-fulfilling prophecies in the classroom. *Journal of Human Capital*, 15(3), 400–431.

Jackson, C.K. (2018) What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy*, 126(5), 2072–2107. Available from: https://doi.org/10.1086/699018

Jacob, B., Dynarski, S., Frank, K. & Schneider, B. (2017) Are expectations alone enough? Estimating the effect of a mandatory college-prep curriculum in Michigan. *Educational Evaluation and Policy Analysis*, 39(2), 333–360. Available from: https://doi.org/10.3102/0162373716685823

Jacob, B.A., Lefgren, L. & Sims, D.P. (2010) The persistence of teacher-induced learning. *Journal of Human Resources*, 45(4), 915–943. Available from: https://doi.org/10.3368/jhr.45.4.915

Kraft, M.A. (2019) Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, 54(1), 1–36. Available from: https://doi.org/10.3368/jhr.54.1.0916.8265r3

Kraft, M.A. (2020) Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. Available from: https://doi.org/10.3102/0013189x20912798

Lavy, V. & Sand, E. (2018) On the origins of gender gaps in human capital: short-and long-term consequences of teachers' biases. *Journal of Public Economics*, 167, 263–279. Available from: https://doi.org/10.1016/j.jpubeco.2018.09.007

Lillard, D.R. & DeCicca, P.P. (2001) Higher standards, more dropouts? Evidence within and across time. *Economics of Education Review*, 20(5), 459–473. Available from: https://doi.org/10.1016/s0272-7757(00)00022-4

Liu, J., Lee, M. & Gershenson, S. (2021) The short-and long-run impacts of secondary school absences. *Journal of Public Economics*, 199, 104441. Available from: https://doi.org/10.1016/j.jpubeco.2021.104441

Liu, J. & Loeb, S. (2021) Engaging teachers measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, 56(2), 343–379. Available from: https://doi.org/10.3368/jhr.56.2.1216-8430r3

Loeb, S. & Candelaria, C.A. (2012) *How stable are value-added estimates across years, subjects and student groups? What we know series: value-added methods and applications. Knowledge Brief 3*. Carnegie Foundation for the Advancement of Teaching.

Mechtenberg, L. (2009) Cheap talk in the classroom: how biased grading at school explains gender differences in achievements, career choices and wages. *The Review of Economic Studies*, 76(4), 1431–1459. Available from: https://doi.org/10.1111/j.1467-937x.2009.00551.x

Murphy, R. & Weinhardt, F. (2020) Top of the class: the importance of ordinal rank. *The Review of Economic Studies*, 87(6), 2777–2826. Available from: https://doi.org/10.1093/restud/rdaa020

Papageorge, N.W., Gershenson, S. & Kang, K.M. (2020) Teacher expectations matter. *The Review of Economics and Statistics*, 102(2), 234–251. Available from: https://doi.org/10.1162/rest_a_00838

Papay, J.P. & Kraft, M.A. (2015) Productivity returns to experience in the teacher labor market: methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105–119. Available from: https://doi.org/10.1016/j.jpubeco.2015.02.008

Papay, J.P., Mantil, A. & Murnane, R.J. (2022) On the threshold: impacts of barely passing high-school exit exams on post-secondary enrollment and completion. *Educational Evaluation and Policy Analysis*, 44(4), 717–733. Available from: https://doi.org/10.3102/01623737221090258

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

Pope, N.G. (2019) The effect of teacher ratings on teacher performance. *Journal of Public Economics*, 172, 84–110. Available from: https://doi.org/10.1016/j.jpubeco.2019.01.001

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

Rivkin, S.G., Hanushek, E.A. & Kain, J.F. (2005) Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. Available from: https://doi.org/10.1111/j.1468-0262.2005.00584.x

Stepner, M. (2013) VAM: Stata module to compute teacher value-added measures. Available from: https://econpapers.repec.org/RePEc:boc:bocode:s457711

Terrier, C. (2020) Boys lag behind: how teachers' gender biases affect student achievement. *Economics of Education Review*, 77, 101981. Available from: https://doi.org/10.1016/j.econedurev.2020.101981

Tran, L. & Gershenson, S. (2021) Experimental estimates of the student attendance production function. *Educational Evaluation and Policy Analysis*, 43(2), 183–199. Available from: https://doi.org/10.3102/0162373720984463

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

Wiswall, M. (2013) The dynamics of teacher quality. *Journal of Public Economics*, 100, 61–78. Available from: https://doi.org/10.1016/j.jpubeco.2013.01.006

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gershenson, S., Holt, S.B. & Tyner, A. (2023) Making the grade: the effect of teacher grading standards on student outcomes. *Contemporary Economic Policy*, 1–14. Available from: https://doi.org/10.1111/coep.12637