



Are female teachers more likely to practice grade inflation? Evidence from China[☆]

Yawen Cheng^a, Dongmin Kong^{b,*}

^a Graduate School of Education, Peking University, China

^b School of Economics, Huazhong University of Science and Technology, China

ARTICLE INFO

Keywords:

Female teachers
Grade inflation
Educational inequality
China

ABSTRACT

Grade inflation has become a global phenomenon that raises concerns in the academia and the public. This study examines whether high grades accompanying with the growth of female teachers arise from student improvement or grade inflation. We use randomly assigned samples from the China Education Panel Survey and obtain the following results. (1) Students taught by female teachers obtain 1.57 grades point higher compared with those taught by male teachers. High grades do not result from student progress but from grade inflation. (2) Female teachers practice grade inflation for different students, and the effect is more pronounced among female students, low-ability students, and students in senior grades. (3) Female teachers may adopt a high-grades practice due to potential pressures under discrimination and uncertainty rather than for the sake of course or students. (4) Last, high grades brought by female teachers lead to less effort for students in the second year. This paper not only provides data support for understanding the increasing grade inflation, but also offers clear implications for the improvement of evaluation systems.

1. Introduction

Regarded as an objective and direct indicator of teaching evaluation, students' grades should convey the student performance and the teacher quality accurately. However, grade inflation has become an inevitable global trend in recent decades.¹ Average grades in most American universities have increased significantly over the past 50 years, from about 2.5 in 1960 to 3.1 in 2006 on a 4-point grading scale (Rojstaczer, 2016). Similar increases in grades have also been observed in Canada, France, and Sweden, among many other countries (Chowdhury, 2018). Such an increase raises concerns on the validity of grading practice and the consequence of grade inflation. Unfortunately, most educators agree that escalation of grades today does not result from improvement of teaching quality or learning skills (Chowdhury, 2018).

[☆] We thank Beijing Normal University for its elaborate organization and arrangement of the seminar. And we thank Feng Hu, Qingyuan Qi, and seminar participants at University of Science and Technology Beijing, Renmin University of China, and Beijing Normal University for helpful suggestions. We acknowledge the financial support of Huazhong University of Science and Technology Double First-Class Funds for Humanities and Social Sciences, and The Fundamental Research Funds for the Central Universities (HUST: 2021WKFZZX007 ; 2022WKFZZX015). All errors are our own.

* Corresponding author.

E-mail addresses: st21133m@gse.pku.edu.cn (Y. Cheng), kongdm@mail.hust.edu.cn (D. Kong).

¹ An increase in grades without a concomitant increase in student quality is known as grade inflation.

<https://doi.org/10.1016/j.chieco.2023.101987>

Received 26 December 2022; Received in revised form 12 April 2023; Accepted 26 April 2023

Available online 11 May 2023

1043-951X/© 2023 Elsevier Inc. All rights reserved.

Grades intend to communicate valuable information. Therefore, grade inflation leads to a range of economic and social consequences for students, schools, and society. In particular, even if lenient grading offers an attractive signal in subsequent academic career and job hunting (Nordin, Heckley, & Gerdtham, 2019), it is highly likely to reduce students' motivation to exert effort in later stages (Griffith & Sovero, 2021). Moreover, grade inflation erodes a school's reputation (Ehlers & Schwager, 2016) and increases the friction of talent matching (Chowdhury, 2018).

However, the potential reasons for grade inflation havenot been well explored. Present studies have mainly used descriptive analysis and mostly focused on grade inflation at higher education levels (Griffith & Sovero, 2021; Jewell & McPherson, 2012; Jewell, McPherson, & Tieslau, 2013; Kolevzon, 1981). Only a few quantitative analyses have been conducted on grade inflation caused by temporary teachers (Chen, Hansen, & Lowe, 2019; Griffith & Sovero, 2021; Ran & Xu, 2019). Accordingly, several mechanisms may contribute to grade inflation. (1) *Grade inflation for different courses*. Structural differences across courses may cause grade inflation, which may lead to markedly lenient grading (Griffith & Sovero, 2021). Some teachers with onerous tasks may resist spending most of their office hours justifying low grades on assignments or examinations, and thereby simply give high grades to avoid possible conflicts and contradictions (Chowdhury, 2018). (2) *Grade inflation for student psychology*. Teachers may also practice grade inflation out of consideration for students. Some educators acknowledge that students now face excessive academic burdens and social pressures, thus teachers may find grade inflation preferable for students' psychological feelings (Finefter-Rosenbluh & Levinson, 2015). (3) *Grade inflation for job reasons*. Teachers play a decisive role in students' grades, however, teacher evaluations largely depend on student performance in turn. That is, teachers may practice grade inflation for job reasons such as job security and financial benefits (Griffith & Sovero, 2021). This phenomenon may be more prominent for female teachers who are faced with considerable discrimination and uncertainty in the labor market (Jewell & McPherson, 2012). For example, women receive significantly lower teaching evaluations from students than their male colleagues (Boring, 2017; Mengel, Sauermaun, & Zölit, 2019). Studies have also found that female faculty members are less likely to receive tenure, recognition, and credit than their male counterparts (Lundberg & Stearns, 2019). Appendix Fig. 1 also shows that compared to male teachers, female teachers are less likely to obtain a high-level title, administrative status, and award.

Mounting evidence has documented that teacher gender plays a non-negligible role in students' grades (Antecol, Eren, & Ozbeklik, 2015; Gong, Lu, & Song, 2018; Lee, Rhee, & Rudolf, 2019). However, few studies have explored the causal relationship between teacher gender and grade inflation, and almost no discussion on grade inflation at secondary education level. To our knowledge, this paper is the first to examine the relationship between teacher gender and grade inflation at secondary education level. The closest literature to us is Griffith and Sovero (2021)'s work. Rather than higher education, the immense benefit of our setting is that the obligatory subject curriculum in secondary education reduces manipulation on the choice of teachers. Moreover, the research on secondary education is important as dynamic formation of human capital dictates that early educational experience may matter more (Cunha & Heckman, 2007). In addition to different learning stage, our research also differs from theirs as follows. First, we conduct more sensitivity analysis to test the cause of high grades, such as using the value-added model and controlling for average scores, etc. Second, we further consider the potential underlying mechanisms under grade inflation, such as course setting, student psychology, and job reasons. Third, we also explore the impact of grade inflation on students' behaviors in the later stage.

Specifically, the randomly assigned samples from the China Education Panel Survey (CEPS, a representative survey on Chinese middle school students) has enabled us to clearly distinguish the impact of teacher gender on students' grades. We find that (1) Students taught by female teachers obtain 1.57 grades point higher compared with those taught by male teachers. High grades are not the result of students learning real knowledge, but the result of grade inflation. (2) Female teachers practice grade inflation for different students, and the effect is more pronounced among female students, low-ability students, and students in senior grades. (3) Female teachers may adopt a high-grades practice due to potential pressures under discrimination and uncertainty rather than for the sake of course or students. (4) Last, high grades brought by female teachers lead to less effort for students in the second year.

Our findings may contribute to present studies in three ways. First, we verify that female teachers have a significant positive impact on students' grades, providing a new factor of grade inflation from the perspective of teacher characteristics (Chowdhury, 2018; Griffith & Sovero, 2021). We also deeply explore the real impact of teacher gender on students' achievements and provide another insight into high grades with female teachers rather than role model in previous studies (Antecol et al., 2015; Gong et al., 2018; Lee et al., 2019). Second, most studies on grade inflation have explored the phenomenon at higher education level (Chen et al., 2019; Griffith & Sovero, 2021; Ran & Xu, 2019). Our paper supplements the literature by indicating that grade inflation may distort educational outcomes and student behaviors from an earlier stage. Third, our setting in China supplements the research on grade inflation of developing countries, and offers us clear policy implications for the global trend of grade inflation. The high proportion of female teachers may be one of the reasons for high scores in recent years. The high grades are not a reflection of real achievements, and may even reduce students' efforts and bring new inequalities. To eliminate pressure-based grade inflation, the government should adopt an open and fair scoring practice, and reduce the possible psychological pressures for female teachers at the source.

The remainder of this paper is organized as follows. Section 2 describes the data and variables. Section 3 provides our conceptual model. Section 4 presents a series of empirical results. Section 5 concludes the study.

2. Data and variables

CEPS² is a large-scale tracking survey project designed and implemented by the China Survey and Data Center (NSRC) of Renmin University of China. Following a stratified, multistage sampling design with probability proportional to size (PPS), approximately 20,000 students in 438 classes of 112 schools across 28 county-level units were randomly selected. CEPS aims to reveal the impact of family, school, and society on individual educational outcomes, providing representative data for relevant academic research and policy implications.

The 2013 baseline survey of students from 7th to 9th grades includes five different questionnaires for students, parents, core subject teachers, head teachers, and school principals. The student questionnaire includes student-level questions, such as test scores, demographic characteristics, and learning habits. The parent questionnaire includes a series of family-level questions, such as parental education and family economic status. And the teacher questionnaire collects teachers' personal information, teaching concepts, and daily work, etc. Above information ideally meets our research needs.³

Among the 112 schools in the baseline survey, 83% ($N = 93$) reported that students are randomly assigned to classrooms upon entry to middle school, and thereby the teachers are randomly assigned to each class. Both students and teachers remain in the same class throughout the three middle school years. Thus the random assignment offers us an appropriate setting to distinguish the impact of teacher gender on students' grades.⁴

In terms of data cleaning and sample selection, we first retain the randomly assigned samples according to the teacher questionnaire.⁵ Second, the teacher data of three core subjects (Chinese, Math and English) are matched to the data of students and parents respectively. Third, we employ standardized scores (with mean of 70 and standard deviation of 10) as teachers' grading⁶ and delete the outliers.⁷ Last, our sample contains 24,388 teacher-student pairs with non-missing necessary variables. Table 1 provides the definitions of main variables in our literature.

Table 2 shows summary statistics for our core variables. To illustrate the role of teacher gender, we present statistical descriptions of 19,006 (78%) female teacher-student pairs and 5382 (22%) male teacher-student pairs respectively. The average scores are 70.79 and 70.26 respectively, with students taught by female teachers perform slightly better than those taught by male teachers. As for student characteristics, gender distribution of students displays a similar pattern in both groups. Schooling years of fathers and mothers in the group taught by female teachers are 10.56 and 9.93 years respectively, while 9.97 and 9.04 years taught by male teachers respectively. As for teacher characteristics, 87% of female teachers are authorized and 84% of male teachers are authorized. And it shows that female teachers are less likely to obtain high-level titles and administrative status compared to male teachers on average.

3. Conceptual model

Specifically, we use the following model to examine the impact of teacher gender on students' grades⁸:

$$Y_{ijgsk} = \alpha + \beta * gender_k + \gamma * X_i + \delta * X_k + \mu_{jgs} + \epsilon_{ijgsk} \quad (1)$$

where Y_{ijgsk} represents the standardized score of subject s at midterm exam in the fall 2013 semester given by teacher k for student i in grade g of school j .⁹ $gender_k$ means the gender of teacher k (1 = female, 0 = male). X_i includes a set of student characteristics, such as student's gender, student's registered residence, whether the student is an only child, schooling years of the student's father and mother, the student's family economic status (1 = very poor, 2 = poor, 3 = medium level, 4 = rich, 5 = very rich) and class size (number of students in student i 's class). Referring to Griffith and Sovero (2021), X_k stands for teacher characteristics of teacher k , such as whether the teacher is authorized (1 = authorized, 0 = otherwise), other classes undertaken by the teacher (number of other classes), the dummy variable indicating the title of the teacher (0 = no, 1 = third-grade teachers, 2 = second-grade teachers, 3 = first-grade teachers, 4 = associate senior teachers, 5 = senior teachers), and the dummy variable indicating the administrative status of the teacher (0 = no, 1 = head of 7th grade or 9th grade, 2 = head of teaching and research office, 3 = academic dean/deputy dean, 4 = vice president, 5 = president). Following the method of Murphy and Weinhardt (2020), we control for the school-by-grade-by-subject μ_{jgs} fixed effect. ϵ_{ijgsk} means the error term. Considering the potential correlation between samples in the same class, standard errors in all

² For detailed information, see <http://ceps.ruc.edu.cn/>.

³ To investigate the real impact of teacher gender, we also employ the second wave of data in our follow-up study.

⁴ We present statistical evidence of the random assignment before formal analysis in Section 4.1.

⁵ The teacher was asked whether the class is divided based on scores of specific subjects. We retain the samples with negative answers.

⁶ The standardized score for each subject is calculated according to the school-grade level respectively.

⁷ We use the "adjacent" command to define the outliers. The "adjacent" command calculates the upper and lower quartiles, p75 and p25, and thus the interquartile range $iqr = p75 - p25$. Among which, adjacent values are the highest value bigger than $p75 + 3/2 iqr$ and the lowest value smaller than $p25 - 3/2 iqr$. It's worth noting that the deleted samples account for only 1.68% and our result still exists with the raw data (see Appendix Table 1).

⁸ We also build a theoretical model to explain the effect of teacher gender, which will be made available on request.

⁹ CEPS only provides students' scores at midterm exam in the fall 2013 semester. Each student has a fixed teacher for each subject, meaning a one-to-one relationship between a student and a teacher.

Table 1
Variable definitions.

Variable	Definition
Core variables	
Grades	Standardized scores for each course
Student characteristics	
Sgender	Student's gender (1 = Male, 0 = Female)
Hukou	Student's registered residence (1 = Non-agricultural, 0 = Agricultural)
Onlychild	Whether the student is only child (1 = Only child, 0 = Otherwise)
Feduy	The schooling years of the student's father
Meduy	The schooling years of the student's mother
Ses	Student's family economic status (1 = Very poor, 2 = Poor, 3 = Medium level, 4 = Rich, 5 = Very rich)
Num	The number of students in the class
Teacher characteristics	
Auth-status	Whether the teacher is authorized (1 = Authorized, 0 = Otherwise)
Multi-class	Other classes undertaken by the teacher
Title	The title of the teacher (0 = No, 1 = Third-grade teachers, 2 = Second-grade teachers, 3 = First-grade teachers, 4 = Associate senior teachers, 5 = Full senior teachers)
Adm-status	The administrative status of the teacher (0 = No, 1 = Head of 7th grade or 9th grade, 2 = Head of Teaching and Research Office, 3 = Academic Dean/Deputy Dean, 4 = Vice President, 5 = President)

Table 2
Descriptive statistics.

Outcome variable	(1)	(2)	(3)	(4)	(5)	(6)
	Female teachers			Male teachers		
	N	mean	sd	N	mean	sd
Grades	19,006	70.790	8.913	5382	70.260	9.288
Student characteristics						
Sgender	19,006	0.500	0.500	5382	0.505	0.500
Hukou	19,006	0.299	0.458	5382	0.273	0.445
Onlychild	19,006	0.507	0.500	5382	0.417	0.493
Feduy	19,006	10.560	3.189	5382	9.972	3.218
Meduy	19,006	9.928	3.455	5382	9.037	3.773
Ses	19,006	2.860	0.568	5382	2.793	0.596
Num	19,006	47.340	12.210	5382	50.580	16.010
Teacher characteristics						
Auth-status	19,006	0.870	0.337	5382	0.840	0.367
Multi-class	19,006	1.021	0.648	5382	1.047	1.000
Title	19,006	2.625	1.083	5382	2.870	1.094
Adm-status	19,006	0.162	0.558	5382	0.646	1.123

Notes: This table displays descriptive statistics for female teachers and male teachers respectively.

regressions are clustered at the class level.¹⁰

With respect to the grading, one may concern that the papers are not graded by teachers, but graded uniformly. Unfortunately, CEPS didn't provide relevant information on who grades the papers, but this issue may not be a big problem in the Chinese context. First, compared with the final exam, the midterm exam is relatively low-stake, generally graded by the teacher who teaches the class. Second, 7th grade test is lower stake than 9th grade test, and we find a similar effect among 7th grade students in subsequent analysis. Finally, the "Comprehensive Evaluation Reform of Elementary and Middle School Education Quality"¹¹ launched by the Ministry of Education in 2013 required that the assessment of compulsory education should not depend on enrollment rate and score ranking, and pointed out that "the tests and surveys should not target at individual students, and it is not allowed to organize unified examinations and tests for all students at or above the county level." In summary, the possibility of unified grading in secondary education is rather small. Another concern lies in whether student scores are comparable. The control for the school-by-grade-by-subject μ_{jgs} fixed effect allows us to compare the effect of teacher gender within the same subject in the same grade and same school. Thus this issue may not

¹⁰ Since a teacher may teach more than one class, we also cluster the standard errors at the school level and the results stay the same, see Appendix Table 2.

¹¹ For more details, see http://www.moe.gov.cn/srcsite/A06/s3321/201306/t20130608_153185.html.

impose a big challenge to our setting.

β comes to our main attention, denoting the average difference in outcomes between the students taught by female teachers and male teachers under the control of other variables.

4. Empirical results

This section first examines the validity of random assignment, provides a series of evidence on grade inflation, and then discusses on possible causes and subsequent consequences of grade inflation.

4.1. Validity of random assignment

The validity of our identification strategy rests on two key assumptions: (1) students are assigned randomly to each subject teacher of our selected sample and (2) students remain with the assigned teacher throughout our analysis sample. Here we verify these assumptions and provide statistical evidence as follows.

First, we identify predetermined variables that are strong predictors of students' grades and possible disturbances of random assignment. Column (1) in [Table 3](#) regresses standardized scores on student characteristics, controlling for the school-by-grade-by-subject fixed effects. The result indicates that student's gender, whether the student is an only child, schooling years of the student's father and mother, and student's family economic status are significant determinants of students' grades.

Given a set of predetermined variables that predict grades, we test the random assignment following the method of [Carrell, Page, and West \(2010\)](#) and [Xu and Li \(2018\)](#). In particular, we regress the teacher's gender on variables that are important predictors of our key outcome measure. If students are randomly assigned to teachers for each subject within each grade and each school, then we expect no association between these characteristics and the assignment to a gender-specific teacher ([Xu & Li, 2018](#)). The result in the column (2) in [Table 3](#) indicates that these significant predictors of students' grades can not predict whether the student is assigned to a female or male teacher. And F-test for the joint significance of these predetermined characteristics also displays a similar pattern. Overall, these tests support the validity of random assignment.¹²

Another concern lies in possible between-class transfers. For example, if a student was randomly assigned to a female teacher in mathematics initially, but he or she strongly preferred a male teacher, his or her parents may transfer the student to another class in which mathematics is taught by a male teacher. However, as students were randomly assigned to three subject teachers upon entry, a transfer to another class means that all teachers may change. Therefore, parents are unlikely to ask for changing due to students' preference for a particular teacher ([Xu & Li, 2018](#)). In addition, our main regression mainly focuses on the first wave, and the possibility of between-class transfer during one year is extremely low.

4.2. Baseline estimation

[Table 4](#) displays the baseline regression. We only include the school-by-grade-by-subject fixed effect in the column (1). The coefficient of $gender_k$ shows a positive correlation between female teachers and students' grades. Column (2) includes all the control variables of student and teacher characteristics, and the coefficient of $gender_k$ remains significant positive. Compared with students taught by male teachers, those taught by female teachers obtain 1.57 grade points higher, equivalent to 2.27% ($1.6 / 70.6$) of the average score.

Our results are comparable to previous research about the effect of adjunct faculty employment on grade inflation. For example, [Chen et al. \(2019\)](#) found that gaining a full-time lecturer position leads to a 0.2–0.3 grade points decline in GPA awarded to the same course by the same instructor, which is nearly equivalent to 7% of the average GPA. [Griffith and Sovero \(2021\)](#) found that female temporary or assistant professors award 0.06 grade points higher than their female tenured colleagues, about 1.88% of the average GPA. Furthermore, it is worth noting that education competition is substantially fierce in China and has gradually concentrated in the early stage. The gap of 1 to 2 points should not be overlooked.

However, whether the high score is the result of student improvement taught by female teachers or grade inflation practice of female teachers remains unclear. Thus, we test these possibilities in the next section.

4.3. Student improvement or grade inflation?

To determine whether the positive impact derives from student improvement or grade inflation, we conduct a series of tests such as employing the value-added model, considering the impact on the entire class, replacing grades with cognitive scores and controlling for the early abilities.

First, we employ the panel feature of CEPS and use the 2013–2014 and 2014–2015 waves of data. If high grades brought by female teachers merely reflect grade inflation (e.g., they do not produce substantial student improvement and just imply lenient grading), then although students obtain high grades in each period, female teachers should not bring about an added value of grades compared with male teachers. Referring to the classical value-added model ([Costello et al., 2008](#); [Rockoff & Cohn, 2015](#)), we add the student's

¹² The random assignment feature of CEPS is also discussed in [Gong et al. \(2018\)](#), [Xu and Li \(2018\)](#), and [Luo, Teng, Xu, and Zhu \(2021\)](#).

Table 3
Balance tests.

Variables	(1)	(2)
	Grades	Teacher gender
Sgender	-3.784*** (0.181)	-0.0001 (0.001)
Onlychild	0.451** (0.205)	-0.003 (0.002)
Feduy	0.300*** (0.038)	-0.001 (0.000)
Meduy	0.114*** (0.031)	-0.00003 (0.0004)
Relatively poor	0.598 (0.455)	0.006 (0.007)
Medium level	0.994** (0.479)	0.012 (0.008)
Relatively rich	0.664 (0.553)	0.006 (0.009)
Very rich	-0.566 (1.486)	0.022 (0.015)
School-by-grade-by-subject fixed effects	Y	Y
Observations	24,388	24,388
Adjusted R-squared	0.080	0.790
F test		1.230 (4.402)

Notes: This table displays balance tests before formal analysis. F test examines the joint significance of the pre-determined variables on teacher gender. Robust standard errors are clustered at the class level. ***p < 0.01, **p < 0.05, *p < 0.1.

previous grades to the right-hand side of Eq. (1) as follows¹³:

$$Y_{ijst} = \alpha + \beta^* gender_k + \theta^* Y_{ijst-1} + \gamma^* X_i + \delta^* X_k + \mu_{js} + \epsilon_{ijst}, \quad (2)$$

where Y_{ijst} represents the standardized score of subject s given by teacher k for student i of school j at year t , $gender_k$ means the gender of teacher k (1 = female, 0 = male), and X_i and X_k stand for student characteristics of student i and teacher characteristics of teacher k respectively. We control for the school-by-subject μ_{js} fixed effect.¹⁴ ϵ_{ijst} is the error term. Standard errors are clustered at the class level. Moreover, β indicates the student's real progress (added value of grades) brought by female teachers.

To begin with, we investigate the impact of teacher gender on the level performance using the model in Eq. (1). Consistent with previous findings, female teachers have a statistically significant and positive effect on the level performance either for students from 7th or 8th grade.

Then, we test whether students improve under the guidance of female teachers. However, there may be concerns that CEPS only tracked students from 7th grade, and there are also some students unsuccessfully tracked.¹⁵ The attrition may not be random and may even relate to teacher gender. If so, our results are probably biased. In this regard, we regress the attrition rate on teacher gender and other control variables on teacher level in the column (3) in Table 5. Specifically, the attrition status takes 1 if the individual student has been successfully tracked in the second wave, and 0 otherwise. The result shows that the attrition does not hinge on teacher gender. Column (4) in Panel A then further examines the impact of teacher gender on the added value of grades between two periods (7th and 8th grade). Similar to Chowdhury (2018), the result manifests that female teachers do not significantly improve students' performances.

Second, if increases in grades merely reflect grade inflation, then we expect high grades for the whole class taught by a female teacher. If this is the case, then the impact of teacher gender on grades should diminish or even disappear after controlling for the mean grades of the student's class given by the teacher. And since female teachers may also care about the degree of the variation of grade distribution, we also control for the variance of class grades. Panel B in Table 5 shows that after controlling for the mean and variance of class grades, students taught by female teachers obtain 0.37 grades point higher, a decrease of nearly 80% compared with the benchmark regression (1.57). The sharp decrease indicates that female teachers may give high grades to the entire class.

Third, given that lenient grades are based on tests in class, teachers should have little or no manipulation on cognitive scores because these tests are uniformly conducted by CEPS project. If female teachers have improved students' achievement, then the positive impact should also reflect on cognitive abilities (Finn et al., 2014). However, the result in Panel C shows that female teachers appear to have no effect on students' achievement.

¹³ Since the second wave (2014–2015) only tracks the students from 7th grade (8th grade in the second wave), thus only the students from 7th grade from 2013 to 2014 wave are preserved.

¹⁴ Since the second wave of survey has only tracked students from 7th grade, we control for the school-by-subject fixed effect.

¹⁵ Owing to attrition, we still employ the data of the first wave for our main analysis.

Table 4
Baseline regression.

Variables	(1)	(2)
	Grades	
Tgender	1.135*** (0.398)	1.571*** (0.336)
Sgender		-3.743*** (0.180)
Hukou		0.951*** (0.210)
Onlychild		0.348* (0.205)
Feduy		0.286*** (0.037)
Meduy		0.100*** (0.031)
Relatively poor		0.608 (0.455)
Medium level		0.982** (0.479)
Relatively rich		0.669 (0.552)
Very rich		-0.663 (1.503)
Num		0.200*** (0.051)
Auth-status		-0.653 (0.643)
Multi-class		0.147 (0.144)
Third-grade teachers		-1.914 (1.252)
Second-grade teachers		-0.564 (0.599)
First-grade teachers		-0.306 (0.638)
Senior teachers		-1.221* (0.721)
Full senior teachers		1.534** (0.667)
Head of 7th grade or 9th grade		0.085 (0.319)
Head of Teaching and Research Office		-0.302 (0.765)
Academic Dean/Deputy Dean		1.117 (0.751)
Vice President		3.046*** (0.939)
School-by-grade-by-subject fixed effects	Y	Y
Observations	24,388	24,388
Adjusted R-squared	0.024	0.085

Notes: This table shows the baseline regression. Robust standard errors are clustered at the class level. ***p < 0.01, **p < 0.05, *p < 0.1.

Last, we also consider the potential effect of students' early abilities on grades at secondary education level, such as score ranking, degree of learning difficulty, and degree of learning diligence in primary school. We can identify the relationship between teacher gender and students' grades more accurately after controlling for these predetermined variables. The almost unchanged result in Panel D in Table 5 again bolsters the confidence in our findings. Furthermore, the main difference between grade inflation and student improvement is that the former comes from teachers' personal behaviors, whereas the latter needs solid efforts of teachers, students and parents. We then also examine the effect on teachers', students' and parents' behaviors and observe no significant impact, the tests will be made available upon request.

Generally, above results suggest that high grades accompanying with female teachers may come from grade inflation rather than student improvement.

4.4. Heterogeneity in teacher gender effect

Having established grade inflation by female teachers, here we turn our attention to how this plays out for different students. These heterogeneity tests not only deepen our understanding on the effect of grade inflation, but also provide us more evidence on the

Table 5
Tests on whether student improvement or grade inflation.

Panel A: Panel data				
Variables	(1)	(2)	(3)	(4)
	Grades in 7th grade	Grades in 8th grade	Tracked successfully	Grades in 8th grade
Tgender	1.908** (0.760)	1.778*** (0.445)	-0.023 (0.058)	0.432 (0.347)
Grades in 7th grade				0.705*** (0.022)
Control variables on student level	Y	Y	Y	Y
Control variables on teacher level	Y	Y	Y	Y
School-by-subject fixed effects	Y	Y	Y	Y
Observations	6885	6885	24,388	6885
Adjusted R-squared	0.101	0.109	0.445	0.512
Panel B: Control for the mean and the variance of class grades				
Variables	(1)	(2)		
		Grades		
Tgender	0.336*** (0.115)	0.371*** (0.135)		
The mean of class grades	Y	Y		
The variance of class grades	Y	Y		
Control variables on student level	N	Y		
Control variables on teacher level	N	Y		
School-by-grade-by-subject fixed effects	Y	Y		
Observations	24,388	24,388		
Adjusted R-squared	0.051	0.105		
Panel C: Cognitive abilities				
Variables	(1)			
	Cognitive abilities			
Tgender	0.072 (0.050)			
Control variables on student level	Y			
Control variables on teacher level	Y			
School-by-grade-by-subject fixed effects	Y			
Observations	24,388			
Adjusted R-squared	0.293			
Panel D: Control for the ability before secondary school				
Variables	(1)			
	Grades			
Tgender	1.098*** (0.351)			
Score ranking	Y			
The degree of learning difficulty	Y			
The degree of learning diligence during primary school	Y			
Control variables on student level	Y			
Control variables on teacher level	Y			
School-by-grade-by-subject fixed effects	Y			
Observations	21,536			
Adjusted R-squared	0.279			

Notes: This table shows whether the higher grades come from student improvement or grade inflation. Column (1), (2) and (4) in Panel A employs two waves of data. Control variables on student level and teacher level are the same with the baseline regression. Unless otherwise specified, the following table stays the same. Robust standard errors are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

robustness of our findings.

First, numerous research has found that boys get lower grades than girls (Falch & Naper, 2013). Also, if the teacher and the student are of the same sex, the student is more likely to get higher scores (Lee et al., 2019). In this regard, we are interested in whether the grade inflation is gender-specific and examine the impact on boys and girls respectively. Panel A in Table 6 shows that female teachers not only give higher grades to girls but also boys. Compared with a male teacher, a female teacher gives 1.12 grades point higher and

Table 6
Heterogeneity in grade inflation.

Panel A: Heterogeneity in gender		
Dependent variable: grades	(1)	(2)
Variables	Boys	Girls
Tgender	1.116** (0.433)	1.795*** (0.371)
Control variables on student level	Y	Y
Control variables on teacher level	Y	Y
School-by-grade-by-subject fixed effects	Y	Y
Observations	12,225	12,163
Adjusted R-squared	0.062	0.067

Panel B: Heterogeneity in ability		
Dependent variable: grades	(1)	(2)
Variables	High-ability	Low-ability
Tgender	1.585*** (0.297)	2.187*** (0.551)
Control variables on student level	Y	Y
Control variables on teacher level	Y	Y
School-by-grade-by-subject fixed effects	Y	Y
Observations	12,194	12,194
Adjusted R-squared	0.096	0.122

Panel C: Heterogeneity in grade		
Dependent variable: grades	(1)	(2)
Variables	7th grade	9th grade
Tgender	1.637*** (0.388)	1.862*** (0.548)
Control variables on student level	Y	Y
Control variables on teacher level	Y	Y
School-by-grade-by-subject fixed effects	Y	Y
Observations	13,235	11,153
Adjusted R-squared	0.089	0.085

Notes: This table shows heterogeneity in student's gender, student's ability and student's grade. Robust standard errors are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7
Grade inflation for different subjects.

Variables	(1)	(2)	(3)
	Chinese	Math	English
Tgender	1.847*** (0.695)	1.645*** (0.426)	2.088* (1.204)
Control variables on student level	Y	Y	Y
Control variables on teacher level	Y	Y	Y
School-by-grade-by-subject fixed effects	Y	Y	Y
Observations	8057	8143	8188
Adjusted R-squared	0.128	0.044	0.128

Notes: This table shows whether the grade inflation appears across different subjects. Robust standard errors are clustered at the class level. ***p < 0.01, **p < 0.05, *p < 0.1.

1.80 grades point higher to boys and girls respectively.¹⁶

Second, although previous studies have found that teachers are highly likely to give high grades to students with good academic performance (Nordin et al., 2019), students with good academic performance can not make substantial progress in grades owing to grades ceiling. Thus, we proxy cognitive scores for students' abilities and estimate our main specification by different abilities separately. The result shows that the teacher gender effect appears in both groups, but differently affects students across different abilities, among which the low-ability group benefits more.

Third, Nordin et al. (2019) found that owing to fierce competitions among schools, grade inflation is highly likely to occur in the senior grade. Therefore, we investigate whether female teachers practice differently to different cohorts. Similar to Nordin et al. (2019), we find a greater effect in the senior grade, with 1.64 grades point higher for 7th grade and 1.86 grades point higher for 9th grade.

Due to students' gender, students' ability and external competition, teachers may conduct different high-grades practice for different students, bringing new inequalities. Above results provide strong support for our findings on grade inflation, but also suggest a double-edged sword especially for those with poor abilities, who may take the high grades for granted and may reduce their efforts in later stages (Griffith & Sovero, 2021).

4.5. Why do female teachers give higher grades?

Moreover, female teachers may practice grade inflation out of course setting, student psychology, and job reasons. Thus, this section attempts to explore the possible mechanisms behind high grades.

First, large class size, additional tasks, and differences in subject may lead to lenient grading (Griffith & Sovero, 2021). Thus, we wonder whether high grades come from different courses taught by female teachers and male teachers. However, grade inflation presents after controlling for class size and other classes undertaken by the teacher in our main analysis. This subsection then further investigates whether the teacher gender effect exists in different subjects. Table 7 verifies that the impact of teacher gender manifests across disciplines, providing us with evidence that the course setting may not be the main reason of grade inflation.

Second, owing to current psychological problems, since female teachers are more sensitive to student psychology, female teachers may practice grade inflation in consideration of student psychology compared with male teachers. Specifically, CEPS asks students "have you had any of the following feelings in the past seven days-dispirited, depressed, unhappy, life is meaningless, sad". The corresponding options are "1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always". We add up these scores and divide the students into one group in a more depressed mood and another in a less depressed mood (the higher the scores, the more depressed is the student). On the contrary, we find that female teachers do not give high scores to the depressed students, but to the group in a less depressed mood in Table 8, suggesting that the grade inflation is not driven by student psychology. And we also employ the student's self-efficacy in 6th grade as a proxy for student psychology and find a similar result, which will be made available upon request.

Third, given the widespread discrimination in the labor market, female teachers may give students high grades out of psychological comfort. Specifically, female faculty members may face greater pressures to issue higher grades, perhaps to counter perceived or actual discrimination, or buffer themselves from potential job insecurity (Jewell & McPherson, 2012). In other words, female academics are more likely to receive less recognition and credit (see Appendix Fig. 1), so that female teachers may find high-grades practice feasible and easy to save the job or even achieve promotion. If so, the effect of teacher gender should be more pronounced in the group that are more stressed with poor student performance. To this end, we use the question "do you usually have a lot of pressures in the following aspects?-students' poor grades" and divide teachers who are under relatively immense pressures with students' grades (answer is under relatively tremendous and tremendous pressures) into one group, and the rest into another group (answer is under normal level, low and no pressures). Panel A in Table 9 shows that the effect of teacher gender seems to only appear in the group with immense pressures with students' grades. Students taught by female teachers obtain 2.23 grades point higher compared with those taught by male teachers in the group with greater pressures.

¹⁶ A reasonable speculation is that female teachers may give female students higher grades out of the same-gender effect (Lee et al., 2019).

Table 8
Grade inflation for student psychology.

Dependent variable: grades	(1)		(2)	
	More depressed		Less depressed	
Variables				
Tgender	1.437***		1.777***	
	(0.418)		(0.452)	
Control variables on student level	Y		Y	
Control variables on teacher level	Y		Y	
School-by-grade-by-subject fixed effects	Y		Y	
Observations	10,453		13,935	
Adjusted R-squared	0.077		0.099	

Notes: This table shows the grade inflation for student psychology. Robust standard errors are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 9
Grade inflation for possible job reasons.

Panel A: Psychological pressure				
Dependent variable: grades	(1)		(2)	
	Under less pressure of students' grades		Under more pressure of students' grades	
Variables				
Tgender	1.448		2.225***	
	(1.284)		(0.481)	
Control variables on student level	Y		Y	
Control variables on teacher level	Y		Y	
School-by-grade-by-subject fixed effects	Y		Y	
Observations	5085		19,303	
Adjusted R-squared	0.103		0.090	

Panel B: Teachers' opinion				
Variables	(1)		(2)	
	Disagree that scores are related to teacher salary and school management		Agree that scores are related to teacher salary and school management	
Tgender	-0.340		1.624***	
	(0.747)		(0.591)	
Control variables of teacher quality	Y		Y	
Control variables on student level	Y		Y	
Control variables on teacher level	Y		Y	
School-by-grade-by-subject fixed effects	Y		Y	
Observations	9460		14,928	
Adjusted R-squared	0.086		0.094	

Notes: This table shows the grade inflation for possible job reasons. Robust standard errors are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 10
Students' response to grade inflation.

Variables	(1)		(2)	
	Confidence degree		Serious attitude	
Tgender	-0.065		-0.139*	
	(0.062)		(0.075)	
Control variables on student level	Y		Y	
Control variables on teacher level	Y		Y	
School-by-grade-by-subject fixed effects	Y		Y	
Observations	6751		6775	
Adjusted R-squared	0.025		0.012	

Notes: This table shows students' response to grade inflation such as confidence degree and serious attitude using two waves of data. Robust standard errors are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Also, if female teachers practice high grades out of job reasons, then the grade inflation may only be observed among teachers who think that students' scores are related to teacher salary and school management. To this regard, we divide the sample to teachers who agree that students' scores are related to teacher salary and school management, and those who disagree according to teacher questionnaire. Consistent with our hypothesis, Panel B in Table 9 shows that the grade inflation is only observed in the group where teachers believe that students' scores are related to teacher salary and school management.

Furthermore, to avoid the disturbances of other teacher characteristics especially for teacher ability, we further control a series of variables such as whether has a qualification certificate, whether is a normal student, years of schooling, years of teaching, whether obtains a "prominent teacher" title and the acquisition of different awards. Appendix Table 3 shows that the result remains still, indicating that teacher quality may not be a principal reason for the grade inflation. Last, we further examine the impact of high scores on future job of female teachers, see appendix for more details.

4.6. How do student respond to female teachers?

Opponents of grade inflation insist that grade inflation may be worrisome to many because students may take the inflated grades for granted, and do not take the study seriously in the later stage. However, some scholars also find it convenient to practice grade inflation, and high grades may even bolster students' confidence in some cases. Therefore, this subsection further investigates the immediate consequence of grade inflation.

Specifically, CEPS asks students "do you feel confident about your future?"¹⁷ and asks parents "overall, what is your child's attitude towards study?"¹⁸ in both waves. We employ the differences of confidence and serious attitude between two periods to proxy for the changes of students. If the coefficient of teacher gender is positive, then it indicates a positive impact on confidence and seriousness, otherwise a negative effect. Column (1) in Table 10 tests whether female teachers enhance students' confidence by higher grades, whereas the column (2) examines whether female teachers have an impact on students' serious attitude. Unfortunately, female teachers seem to reduce students' serious attitude on study, but not boost their confidence.

5. Conclusions

Using the randomly assigned samples from CEPS, we find that female teachers are more likely to practice grade inflation than their male counterparts. And the effect is more pronounced among female students, low-ability students and senior grade students. Moreover, the high-grades practice results from possible job reasons instead of course setting and student psychology, reducing the students' effort in the later stage.

These findings provoke our thought on the possible cause of high grades and enhance our understanding of grade inflation. Specifically, female teachers may practice grade inflation out of psychological pressures for possible job reasons neither for the sake of different course or student psychology. These results provide suggestive evidence with the limited literature that finds job insecurity induces instructors to increase grades, especially for female teachers. Furthermore, the effect of grade inflation may lead to less effort in the later stage, resulting in new inequalities and distortions to different students. To avoid educational inequalities and mismatches in the future, the government should adopt an open and fair scoring practice on the one hand, and further reduce the possible psychological pressures for female teachers from the origin.

However, note that due to data restriction, we can not offer direct evidence on why female teachers intend to give high grades. And with two waves of data, we also can not examine either the long-term effect of grade inflation on students or whether teachers actually get promoted after high-grades practice in the long run. Future research is expected to explore other hidden factors and long-term consequences of grade inflation.

Declaration of Competing Interest

We, the authors of this paper, declare that: 1) we have no conflict of interest that could have appeared to influence the work reported in this paper, and 2) we have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

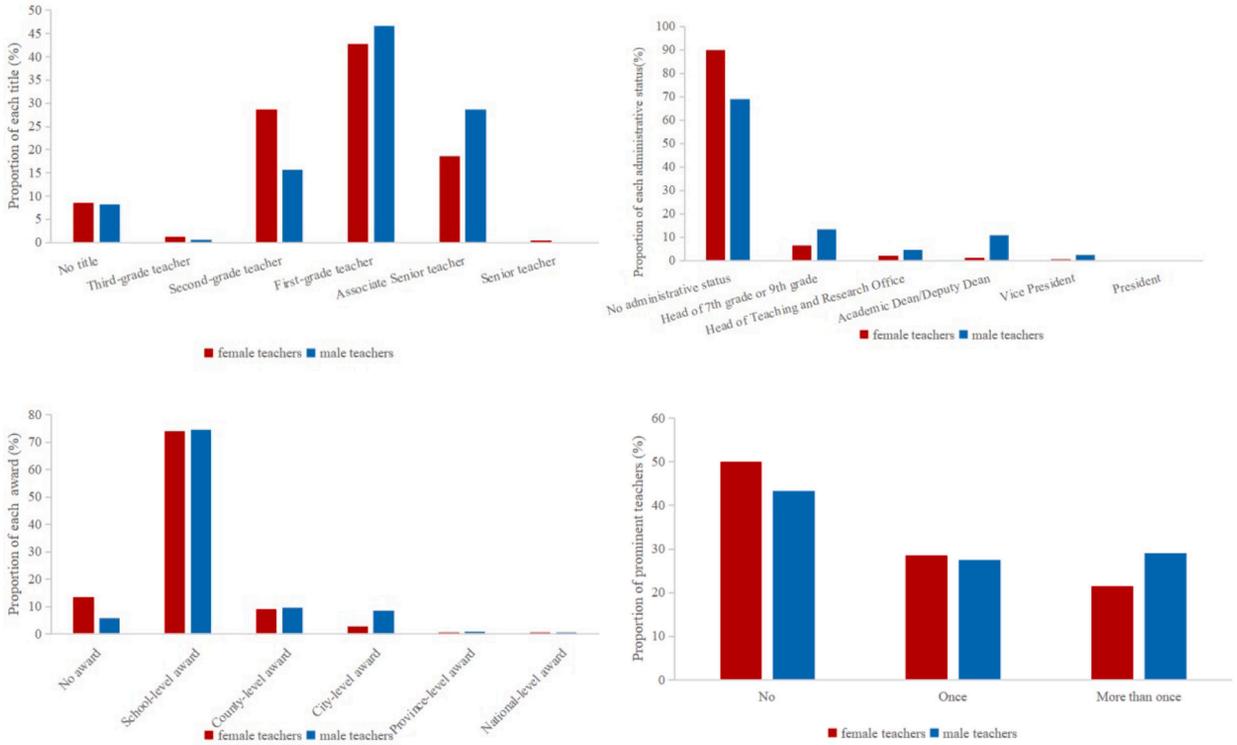
Data availability

Data will be made available on request.

Appendix A. Appendix

¹⁷ The answer is 1–4 ratings, 1 = not confident at all, 2 = not very confident, 3 = a little confident, 4 = very confident. The higher the rating, the more confident is the student in the future.

¹⁸ The answer is 1–5 ratings, 1 = not serious at all, 2 = not very serious, 3 = a little serious, 4 = very serious. The higher the rating, the more serious is the student about the study.



Appendix Fig. 1. Titles and awards for male and female teachers.

Appendix Table 1

Grade inflation with raw data.

Variables	(1) Grades
Tgender	1.475*** (0.336)
Control variables on student level	Y
Control variables on teacher level	Y
School-by-grade-by-subject fixed effects	Y
Observations	24,804
Adjusted R-squared	0.088

Notes: This table shows grade inflation with raw data. Robust standard errors are clustered at the class level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Table 2

Grade inflation with alternative standard error.

Variables	(1) Grades
Tgender	1.571*** (0.457)
Control variables on student level	Y
Control variables on teacher level	Y
School-by-grade-by-subject fixed effects	Y
Observations	24,388
Adjusted R-squared	0.085

Notes: This table shows grade inflation with standard errors clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Table 3
Grade inflation controlling for teacher quality.

	(1)
Variables	Grades
Tgender	1.224*** (0.362)
Control variables of teacher quality	Y
Control variables on student level	Y
Control variables on teacher level	Y
School-by-grade-by-subject fixed effects	Y
Observations	23,822
Adjusted R-squared	0.086

Notes: This table shows grade inflation controlling for teacher quality. ***p < 0.01, **p < 0.05, *p < 0.1.

Higher grades may be correlated with more favorable outcomes for female teachers. Referring to Griffith and Sovero (2021), we use the panel data feature in investigating the title upgrading¹⁹ among teachers. Taking the job promotion as the explained variable (1 = the teacher gets promoted compared to the first wave, 0 = otherwise), we take the sample of teacher as unit *i*. The core explanatory variables turn to the gender of teacher *i*, average grades taught by teacher *i* at year *t* - 1, and the interaction term of these two variables. The coefficient of the interaction term δ is of our main focus, reflecting whether female teachers experience a significant higher probability of being promoted for higher grades compared with their male counterparts.

$$Y_{ijst} = \alpha + \beta * gender_i + \delta * gender_i * average_grades_{ijst-1} + \theta * average_grades_{ijst-1} + \gamma * X_i + \mu_{js} + \epsilon_{ijst} \tag{3}$$

Y_{ijst} represents the job promotion of subject *s* for teacher *i* in school *j* at year *t*, $gender_i$ means the gender of teacher *i* (1 = female, 0 = male), $average_grades_{ijst-1}$ indicates the average grades of subject *s* given by teacher *i* in school *j* at year *t*-1, and X_i stands for teacher characteristics of teacher *i* as the baseline model. Moreover, we control for the school-by-subject μ_{js} fixed effect, and ϵ_{ijst} means the error term. Standard errors are clustered at the class level.

Consistent with our hypothesis, female teachers have a higher probability of being promoted for high grades compared with their male colleagues. This result may be the case why female teachers are more likely to practice grade inflation. However, due to data restriction, above results need to be interpreted carefully. Specifically, as we do not have direct information about the real motivation of grade inflation, we provide these tests as complementary evidence as well as a suggestion for future work.

Appendix Table 4
Title upgrading.

	(1)
Variables	Title upgrading
Tgender * average grades of teacher	0.475* (0.276)
Control variables on teacher level	Y
School-by-subject fixed effects	Y
Observations	169
Adjusted R-squared	0.267

Notes: This table employs teacher-level data and shows the grade inflation for possible job reasons. Robust standard errors are clustered at the class level. ***p < 0.01, **p < 0.05, *p < 0.1.

References

Antecol, H., Eren, O., & Ozbeklik, S. (2015). The effect of teacher gender on student achievement in primary school. *Journal of Labor Economics*, 33(1), 63–89.
 Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27–41.
 Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics*, 125(3), 1101–1144.
 Chen, K., Hansen, Z., & Lowe, S. (2019). Why do we inflate grades? The effect of adjunct faculty employment on instructor grading standards. *The Journal of Human Resources*, 56(3), 878–921.
 Chowdhury, F. (2018). Grade inflation: causes, consequences and cure. *Journal of Education and Learning*, 7(6), 86–92.
 Costello, R., Elson, P., Schacter, J., Doctors, T., Network, V. A., & Delta, A. P. (2008). An introduction to value-added analysis. *Catholic Education A Journal of Inquiry & Practice*, 12, 193–203.

¹⁹ The title of the teacher (0 = No, 1 = Third-grade teachers, 2 = Second-grade teachers, 3 = First-grade teachers, 4 = Associate senior teachers, 5 = Full senior teachers).

- Cunha, F., & Heckman, J. (2007). The technology of skill formation. *The American Economic Review*, 97(2), 31–47.
- Ehlers, T., & Schwager, R. (2016). Honest grading, grade inflation, and reputation. *CEISifo Economic Studies*, 62(3), 506–521.
- Falch, T., & Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review*, 36, 12–25.
- Finefter-Rosenbluh, I., & Levinson, M. (2015). What is wrong with grade inflation (if anything)? *Philosophical Inquiry in Education*, 23(1), 3–21.
- Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., ... Gabrieli, J. D. E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science*, 25, 736–744.
- Gong, J., Lu, Y., & Song, H. (2018). The effect of teacher gender on students' academic and noncognitive outcomes. *Journal of Labor Economics*, 36(3), 743–778.
- Griffith, A. L., & Sovero, V. (2021). Under pressure: How faculty gender and contract uncertainty impact students' grades. *Economics of Education Review*, 83, Article 102126.
- Jewell, R. T., & McPherson, M. A. (2012). Instructor-specific grade inflation: Incentives, gender, and ethnicity. *Social Science Quarterly*, 93(1), 95–109.
- Jewell, R. T., McPherson, M. A., & Tieslau, M. A. (2013). Whose fault is it? Assigning blame for grade inflation in higher education. *Applied Economics*, 45(9), 1185–1200.
- Kolevzon, M. S. (1981). Grade inflation in higher education: a comparative study. *Research in Higher Education*, 15(3), 195–212.
- Lee, J., Rhee, D. E., & Rudolf, R. (2019). Teacher gender, student gender, and primary school achievement: evidence from ten Francophone African countries. *Journal of Development Studies*, 55(4), 661–679.
- Lundberg, S., & Stearns, J. (2019). Women in economics: stalled progress. *The Journal of Economic Perspectives*, 33(1), 3–22.
- Luo, R., Teng, K., Xu, L., & Zhu, Y. (2021). A decomposition of student achievement gap by gender in China: evidence from random class assignment. *International Journal of Educational Research*, 106, Article 101721.
- Mengel, F., Sauermaun, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535–566.
- Murphy, R., & Weinhardt, F. (2020). Top of the class: the importance of ordinal rank. *The Review of Economic Studies*, 87(6), 2777–2826.
- Nordin, M., Heckley, G., & Gerdtham, U. (2019). The impact of grade inflation on higher education enrolment and earnings. *Economics of Education Review*, 73, Article 101936.
- Ran, & Xu, R. D. (2019). Does contractual form matter?: The impact of different types of non-tenure-track faculty on college students' academic outcomes. *The Journal of Human Resources*, 54.
- Rockoff, J. E., & Cohn, E. (2015). Value-added modeling: a review. *Economics of Education Review*, 47, 180–195.
- Rojstaczer, S. (2016). Grade inflation at American colleges and universities. www.gradeinflation.com Accessed November 4, 2016.
- Xu, D., & Li, Q. (2018). Gender achievement gaps among Chinese middle school students and the role of teachers' gender. *Economics of Education Review*, 67, 82–93.