Title
Essays on the Economics of Education

Author
Paredes Haz, Valentina Amanda

Publication Date
2012

Peer reviewed|Thesis/dissertation

Essays on the Economics of Education

By

Valentina Amanda Paredes Haz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Card, Chair

Professor Patrick Kline

Professor Steven Tadelis

Spring 2012

I would like to thank my adviser, David Card, for his support and guidance over the course of my graduate studies.

I would also like to thank Patrick Kline, Steven Tadelis, Enrico Moretti, Shachar Kariv, Issi Romem, Edson Severnini, Monica Deza, Willa Friedman, Francois Gerard, Rodolfo Lauterbach and Ricardo Paredes for their comments and help. All errors are my own.

Contents

1. Grading System and Student Effort 1
1.1. Introduction 1
1.2. Related Literature 2
1.3. Model 4
1.3.1. Absolute Grading 4
1.3.2. Relative Grading 6
1.3.3. Comparing the Two Regimes 8
1.4. Empirical Strategy 12
1.5. Data 14
1.6. Results 16
1.6.1. Structural Estimation Results 19
1.6.2. Policy Simulations 19
1.7. Conclusions 24

2. Grading Standards and Student Achievement 25
2.1. Introduction 25
2.2. Data 26
2.3. Theoretical Framework 27
2.3.1. Effect of Grading Standards on Student Effort 27
2.3.2. Estimating the Effect 29
2.4. Results 31
2.4.1. Estimating the Standards 31
2.4.2. Effect on Test Scores 32
2.5. Specification Checks 34
2.5.1. Differences in Curriculum Coverage 34
2.5.2. Nonlinearities in the Effect 35
2.5.3. Different Effect by Subject 35
2.6. Conclusions 37

3. Does teacher gender matter? 38
3.1. Introduction 38
3.2. Data 39
3.3. Econometric Framework 40
3.4. Results 42
3.5. Robustness Checks 43
3.6. Possible Mechanisms 47
3.7. Conclusions 49

A. Appendix: Proofs of Results 53

List of Tables

1. Absolute versus Relative Grading Systems 15
2. Grading System, Entry Score, Percentage Male and Passing Rate by Period 16
3. Effect of Relative Grading on Students' Grades: Aggregate Data 17
4. Effect of Relative Grading on Students' Grades: Individual Data 17
5. Effect of Relative Grading on Students' Grades: Upper and Lower Bounds 19
6. Structural Parameters 21
7. Model Fit: Actual versus Simulated Moments 22
8. Descriptive Statistics by Section: No Tracking 23
9. Effect of Tracking on Grades 23
10. Descriptive Statistics 27
11. Estimated Standards 31
12. Correlation of Estimated Standards 31
13. Standards and School Characteristics 32
14. Estimated effects of grading standards on test scores 32
15. Estimated effects of grading standards on test scores: Quantile regressions 34
16. Estimated effects of grading standards on test scores controlling by curriculum 35
17. Estimated effects of grading standards on test scores: Exploring nonlinearities 36
18. Estimated effects of grading standards on test scores: Effect by Subject . 36
19. Gender Gap in SIMCE 2009: Scores by Subject 39
20. Data: Sample versus Population 40
21. Gender Gaps by Subject 41
22. Teacher Characteristics by Gender 41
23. Estimated effects of female teachers on test scores 43
24. Estimated effects of female teachers on test scores: Teacher Fixed Effects 44
25. Estimated effects of female teachers on test scores: Classroom Fixed Effects 45
26. Probability of Being Assigned to a Female Teacher 45
27. Estimated effects of female teachers on test scores 46
28. Percentage of Female Teachers by Subject 47
29. Estimated effects of female teachers on test scores by subject 48
30. Estimated effects of female teachers on test scores by mother's level of schooling 49

List of Figures

1. Effort by Grading System 10
2. Effort as a Function of Uncertainty 11
3. Effect of Relative Grading 18
4. Grade Kernel Densities 20
5. Effect of Relative Grading 21
6. Grade Distribution 28
7. Optimal Effort 28
8. Effect of an Increase in Standards 29

Abstract

Essays on the Economics of Education

by

Valentina Amanda Paredes Haz

Doctor of Philosophy in Economics

University of California, Berkeley

Professor David Card, Chair

An important question that many educators face is how to motivate students to study. Many programs in the US and other countries give cash or award incentives to encourage students to exert more effort. In the following three essays, I explore different alternatives to raise student effort, which in turn should raise student achievement, measured in grades and standardized test scores.

In my first essay, I propose that the grading system affects the incentives to exert effort among students. For this purpose, I build a model where students maximize their utility by choosing effort. I investigate how student effort changes when there is a change in the grading system from absolute grading to relative grading. I use data from college students in Chile who faced a change in the grading system to test the implications of my model. My model predicts that, for low levels of uncertainty: (i) total effort is higher with absolute grading; (ii) low ability students exert less effort with absolute grading, and; (iii) high ability students exert more effort with absolute grading. The data confirms that there is a change in the distribution of effort, although I don't find a change in the total level of effort.

One results from the model discussed in the first essay is that high ability students exert higher effort under higher standards, but a high standard might have a negative impact on low ability students, who could give up and hence exert zero effort. So in my second essay, I explore whether higher grading standards have an effect on student achievement measured by standardized tests. Grading standards are measured as the school intercept in a regression of standardized test scores on grades. Using data from 8th graders in Chile, I find that higher standards have a positive average effect on standardized test scores. This effect is positive for the percentiles 25, 50 and 75 of the achievement distribution and is larger for the 25th percentile.

In my third essay, I explore whether the gender of the teacher has an impact on student achievement and if this impact is different for boys and girls. Again, I use data from 8th graders in Chile. Within-student comparisons based on these data indicate that assignment to a same-gender teacher significantly improves the achievement of girls but doesn't improve the achievement of boys. I find that the effect is larger for subjects that are traditionally considered male dominated, and for girls whose mothers have low levels of education, which is consistent with a role model hypothesis.

1. Grading System and Student Effort

1.1. Introduction

There is interest in scientific and public debates about ways to induce more effort from students. Large amounts of money are being spent to give incentives to students, especially students from minority groups, to exert more effort. In a recent survey of reward programs, Raymond (2008) finds that different districts, including Baltimore, Maryland, Fulton County, Georgia and New York City, have programs for K-12 public school students that range from giving cash or MP3 players to students, to rewards such as social events or concerts.

In this debate, little attention is devoted to the possible influence of the type of grading system on student effort. This essay studies how students respond to grading incentives. In particular, I want to answer two questions: 1) How is total effort affected by the grading system? and, 2) How is the distribution of effort affected by the grading system? Compared to reward programs, changing the grading system is both cheaper and less controversial.

I study and compare two grading systems widely used in schools and universities: relative grading and absolute grading. I classify a grading system as absolute when students pass the class if they meet a fixed standard that is given ex ante. On the other hand, I classify a system as relative when the standard depends on the performance of the students in the class. Grading on a curve is then considered relative grading.

The first part of the essay models the student decision on how much effort to devote to learning. I model student effort under the two grading systems and compare the total expected effort and the distribution of effort. The model helps to understand how students' incentives to exert effort change under the two grading regimes.

The second part tests the implications of the theoretical model using data of students' grades from a university in Chile where the grading system changed from absolute to relative. In my data set I don't observe effort directly, but I can observe it indirectly through grades. I thus use the grade distribution before and after the change in the grading system to identify the causal impact of the new grading system.

My model predicts that, for low levels of uncertainty: (i) total effort is higher with absolute grading; (ii) low ability students exert less effort with absolute grading, and; (iii) high ability students exert more effort with absolute grading. The data confirms that there is a change in the distribution of effort, although I don't find a change in the total level of effort.

This study contributes to the literature in two ways. First, it studies the distributional effects of a change from an absolute standard to a relative standard. This is relevant in the context of grading, but is also relevant in personnel economics. Second, my study provides empirical evidence to the predictions of the theoretical model. I believe this is the first study that uses a change in the grading regime to test the consequences of the grading regime on students' effort and empirically compares these two grading systems.

The remainder of this discussion is structured as follows. Section 1.2 discusses the relevant literature. Section 1.3 develops the model. Section 1.4 describes the empirical strategy. Section 1.5 describes the data. Section 1.6 presents the results and Section 1.7 concludes.

1.2. Related Literature

There has been some research on how student effort responds to cash incentives. Angrist and Lavy (2009) study how cash incentives increase the Bagrut certification rates among low achieving students, using experimental data. The experiment uses a school-based randomization design offering rewards to all students in treated schools who passed their exams. They found an increase in certification rates for girls on the order of 0.10 but they found no significant effect for boys.

Angrist, Lang and Oreopoulos (2009) report on a randomized field experiment involving two strategies designed to improve achievement among first year undergraduates at a large Canadian University. One treatment group was offered peer advising and organized study group services. Another treatment group was offered substantial merit-scholarships equivalent to a full year's tuition. A third group was offered both interventions. They found that no program had an effect on grades for males; however, first-term grades were significantly higher for females in the two scholarship treatment groups.

Kremer, Miguel and Thornton (2009) examine the impact of a merit program for adolescent girls introduced in rural Kenya primary schools. Girls who scored well on academic exams received a cash grant and had school fees paid. Girls eligible for scholarships showed significant gains in academic exam scores, with an average gain of 0.15 standard deviations. They also found evidence of positive externalities on learning: boys, who were ineligible for the awards, also showed sizeable average test gains, as did girls with low pretest scores. Both teacher and student school attendance increased in the program schools.

There are a few papers which study the effect of grading in student effort. Kang (1985) compares a pass-fail reward system with a fixed standard and rewards for improvements in performance. He shows that, when a pass-fail system is used, raising the passing standard increases effort up to a point. Beyond this point, however, students start giving up because the effort required to pass the class is more costly than the expected reward. Costrell (1994) studies the determination of standards for educational credentials, such as high school diplomas. His model has three features: (a) the standard is defined as the required level of proficiency for a binary credential; (b) the standard is set by policymakers who maximize their conception of social welfare, and; (c) utility-maximizing students choose whether to meet the standard. He shows that more egalitarian policymakers set lower standards, that the median voter would prefer higher standards, and that decentralization lowers standards.

Betts (1998) also models the determination of standards. The main difference between his paper and Costrell's is that students differ in ability. He finds that an egalitarian policymaker might prefer higher standards than would a policymaker whose goal was to maximize the sum of earnings. The result is based on the fact that an increase in educational standards increases earnings of both the most able and the least able workers.

Betts and Grogger (2003) study the distributional effects of increasing grading standards. They find that high standards increase the test scores for all students but that the increase is greatest for the top of the test distribution. They also find no effect on educational attainment and a negative effect on high school graduation for blacks and Hispanics.

Figlio and Lucas (2004) study the effects of teacher-level grading standards on student achievement. They also find that the students who benefit most from higher standards are the initially high achieving students. In addition, they find that low achieving students benefit more from high standards when their classmates are high achieving and that high achieving students benefit more from higher standards when their classmates are low achieving.

Babcock (2009) uses course evaluations from UC San Diego to see how average study time is affected by grading standards. He finds that average study time would be about 50% lower in a class in which the average expected grade was an A versus a class in which the average expected grade was a C.

Using the tournament model developed by Lazear and Rosen (1981), Becker and Rosen (1992) compare relative to absolute grading. They assume that students are risk neutral and that they can get a reward of W_1 if they pass the class and a reward of W_2 if they fail the class. Scores are produced by effort and by a random component that can be divided into a common shock and an idiosyncratic shock. They argue that competition between students does stimulate academic effort, provided students are appropriately rewarded for achieving. They also argue that stratifying students into groups in which each has a chance for success may be preferred to the setting of a single national standard.

Another paper that compares relative and absolute grading is Dubey and Geanakoplos (2010). They show that absolute grading is always better than grading on a curve. The main difference between their model and mine is that, in their model, students care only about their relative ranking, while, in my model, students only care about passing or failing the class.

My study is also related to the contest literature, in particular to the early literature that compares tournaments to individual contracts such as piece-rates and absolute standards. An early paper that compares a fixed standard against a relative standard is Lazear and Rosen (1981). Their paper models promotions as a relative game, where relative performance is a way of generating incentives. Lazear and Rosen compare three compensation schemes: a linear piece rate, a fixed standard, and a tournament. In the first, each agent's compensation is a linear function of his output; in the second, each agent receives one of two fixed payments, depending on whether his output is above or below a specified standard; and in the third, the agents compete against each other for two fixed prizes, allocated on the basis of the rank order of their output levels. They find that, if workers are risk neutral, all schemes are optimum labor contracts. However, when risk aversion is introduced, the prize salary scheme no longer duplicates the allocation of resources induced by the optimal piece rate. The intuition is that a tournament is a nonlinear transformation of output to wages, so the expected income and variance differ in the two schemes. Lazear and Rosen are not able to completely characterize the conditions under which piece rates dominate rank-order tournaments. Green and Stokey (1983) show that, when agents are risk averse and the principal is risk neutral, and in the absence of a common shock, using optimal independent contracts dominates using the optimal tournament. Conversely, if the distribution of the common shock is sufficiently diffuse, using the optimal tournament dominates using optimal independent contracts. Green and Stokey assume ex ante identical agents, so they don't have to consider the distribution of effort.

Compared to personnel economics, where there is a principal who wants to maximize the expected output minus payment, the context of education differs because it is not clear what the objective function of the teacher should be. This is why, aside from focusing on total effort, I also focus on the distribution of effort.

My model also differs from this previous literature because I assume that the teacher is not able to affect the student's valuation of passing the class, while in the case of promotion studied by Lazear and Rosen (1981), the principal can affect the valuation of the agent getting a promotion by changing the wage structure.

1.3. Model

A teacher wants to induce N students to exert effort through the allocation of grades. The allocation depends on a score S_i, which is an average of J partial scores, S_{ijk}, where i denotes the student, j denotes the test and k denotes the teacher. Each partial score is a function of effort and a shock, $S_{ijk} = e_i + \varepsilon_{ijk}$[1], such that $S_i = e_i + \frac{\sum_{j=1}^{J} \varepsilon_{ijk}}{J}$. Student i gets utility $U_i(Success)$ if he passes the class and $U_i(Fail)$ if he doesn't. Notice that students in this model don't care about grades per se, but they care about passing the class[2]. The utility of student i of passing the class by exerting effort e_i is $u_i^* = U_i(Success) - c_i e_i$. Utility is normalized such that $U_i(Fail) = 0$. Students can exert effort $e_i \in [0, \bar{e}]$, where $\bar{e}$ is such that $U_i(Success) - c_i\bar{e} < 0$. This assumption means that is never optimal for the student to devote all his available time to studying.

Because behavior is invariant to affine transformations, I rewrite the utility function as $u_i \equiv \frac{u_i^*}{c_i} = v_i - e_i$, where $v_i = \frac{U_i(Success)}{c_i}$. In what follows, I will refer to v_i as ability and will interpret a higher v_i as higher ability because the cost of exerting effort is lower.

I assume that $U_i(Success) = U(Success)$ so students differ only in their cost of effort c_i. Let $c_1 < c_2 < c_3 < ... < c_N$ so $v_1 > v_2 > ... > v_N$.

I will use the following assumption

Assumption 1.1 $\frac{\sum_{j=1}^{J} \varepsilon_{ijk}}{J} \to \theta_k$

This assumption can be interpreted as there being a large number of tests and $\varepsilon_{ijk} = \eta_{ij} + \theta_k$, where the first term is i.i.d and so the law of large numbers applies, but the second term, which can be interpreted as a teacher or classroom level shock, remains. Therefore, from now on, I will use the following production function $S_i = e_i + \theta_k$

Let $\theta_k \sim G$, where G is a symmetric, twice differentiable cumulative distribution function with $E[\theta] = 0$, $Var(\theta) = \sigma_\theta^2$, and pdf g.

1.3.1. Absolute Grading

In the case of absolute grades, students pass the class if $S_i \geq S^*$ and fail otherwise, i.e. students pass the class if their score is above a certain threshold, where S^* is exogenously

[1] To make the analysis simple, I assume that students cannot change their effort for different tests.

[2] Another interpretation is that students do care about grades, but the discontinuity in utility from passing versus failing is sufficiently large that we can simplify the model by saying that students only care about passing.

given. Students solve the following optimization problem:

$$\max_{e_i} \quad v_i Pr(S_i \geq S^*) - e_i \qquad s.t. \qquad S_i = e_i + \theta$$
$$\Rightarrow \max_{e_i} \quad v_i G(\frac{e_i - S^*}{\sigma_\theta}) - e_i$$

The FOC is

$$\begin{aligned} \Rightarrow \quad \frac{v_i}{\sigma_\theta} g(\frac{e_i^* - S^*}{\sigma_\theta}) - 1 &= 0 \quad and \quad e_i^* > 0, \quad or \\ \frac{v_i}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta}) - 1 &\leq 0 \quad and \quad e_i^* = 0 \end{aligned}$$

and the SOC

$$\frac{v_i}{\sigma_\theta^2} g'(\frac{e_i^* - S^*}{\sigma_\theta})$$

so e_i^* is a local maximum if and only if $SOC < 0$. Suppose there is an interior solution to the maximization problem. Then

$$e_i^* \quad = \quad S^* + \sigma_\theta g^{-1}(\frac{\sigma_\theta}{v_i})$$

Let $e(v)$ denote the effort function. Then

$$e(v_i) = \begin{cases} S^* + \sigma_\theta x_i & \text{if } v_i G(x_i) - S^* - \sigma_\theta x_i \geq v_i G(\frac{-S^*}{\sigma_\theta}); \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

where $x_i(\sigma_\theta) = g^{-1}(\frac{\sigma_\theta}{v_i})$

Let K be the expected number of students who pass the class. If the teacher wants only K students to pass the class, she must set the threshold S^* such that the marginal student (student K) is indifferent between actively participating and exerting no effort. In other words, she must set S^* such that

$$v_K G(x_K) - S^* - \sigma_\theta x_K = v_K G(\frac{-S^*}{\sigma_\theta})$$

It is useful to see what happens to effort when there is no common shock, i.e. $\sigma_\theta = 0$. Then

$$e(v_i) = \begin{cases} S^* & \text{if } v_i \geq S^*; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

and total expected effort is given by

$$W^A = \mathbb{E}[\#Passes]S^* \tag{3}$$

Because the teacher wants only K students to pass the class, she sets $S^* = v_k$ and total expected effort can be rewritten as

$$W^A = Kv_K \tag{4}$$

When there is imperfect information, the teacher should set $S^* = \int_0^{\bar{v}} v_K dF_K(v_K)$. Total expected effort is then

$$W^A = K \int_0^{\bar{v}} v_K dF_K(v_K) \tag{5}$$

The following three results establish what happens to the threshold and effort as the variance of the common shock, σ_θ, goes up.

Result 1.2 *If the variance of the common shock increases, then the standard needs to decrease such that K prizes are given in expectation.*

Proof See the Appendix

Result 1.3 *The effort of the marginal student is decreasing on the variance of the shock.*

Proof See the Appendix

Result 1.4 *When $v_i - v_K$ is large enough, the effort of student i can be increasing on the variance of the shock.*

Proof See the Appendix

When the number of students in a class, N, increases, but the teacher wants to leave K constant, total expected effort increases. The intuition for this is as follows: when N increases, there are more students of high ability, so the standard must increase to keep K constant. Then, the K students who pass the class are also exerting more effort.

Increasing the number of students in a class also has an effect on the distribution of effort. Because the standard has to increase to leave K constant, high ability students exert more effort. However, there will be some students who were exerting a positive amount of effort before, but, with the increase in S^*, now exert no effort.

When the teacher wants to increase K, keeping the number of students in the class constant, total expected effort may increase or decrease. To increase K, the teacher must lower the standard. Then, there are more students exerting a positive amount of effort, but they are exerting less effort than before.

Because the increase in K causes a decrease in S^*, high ability students exert less effort. However, there will be some students who were exerting zero effort before and now exert a positive amount of effort.

1.3.2. Relative Grading

When grades are relative, instead of fixing the threshold S^*, the teacher fixes the number of students who pass the class. I use the following notation: N is the total number of students and K is the total number of passes. Relative grading can be modeled as an all pay auction of K homogenous goods[3]. Notice that the common shock doesn't affect students' behavior because only rank matters and the shock is rank preserving. In other words, relative grading provides an insurance against the common shock, θ_k.

[3] The teacher is auctioning several prizes to the students, where the prize is to pass the class. The students compete for the prizes by producing a score through effort. It is all-pay because, once effort is exerted, the student cannot recover his cost of investing in effort, even if he fails the class

Suppose there are N students, with valuations $v_1 > v_2 > ... > v_N$. At the time of exerting effort, each student knows N, K, and $v_1, v_2, ..., v_N$ (valuations are common knowledge).

It can be shown that there is no equilibrium in pure strategies in this game[4]. Let $H_i(e)$ represent the cumulative density function of student i's equilibrium mixed strategy, with support $[v_i^l, v_i^u]$. The unique equilibrium in mixed strategies is described in Clark and Riis (1998). In this equilibrium, the K+1 students with highest valuation participate actively. Students 1 to K exert a strictly positive amount of effort while student K+1 randomizes between zero and a positive amount of effort[5]. The remaining students exert zero effort with probability of one. The expected utility of student i is $v_i - v_{K+1}$ if $v_i \geq v_{K+1}$ and zero otherwise. The expected amount of effort is the following

$$\mathbb{E}[e_i] = \begin{cases} 0 & \text{if } v_i < v_{K+1} \\ P_i v_i - (v_i - v_{K+1}) & \text{if } v_i \geq v_{K+1} \end{cases} \tag{6}$$

where P_i is the probability that student i passes the class, conditional on his level of effort and the level of effort of the other N-1 students in a Nash equilibrium.

Notice that $\mathbb{E}[e_i] = v_{K+1} - (1 - P_i)v_i < v_{K+1} \leq v_i$ for $v_i \geq v_{K+1}$

Total expected effort is given by

$$W^R = \sum_{i=1}^{N} \mathbb{E}[e_i] = \sum_{i=1}^{K+1} [P_i v_i - (v_i - v_{K+1})] \tag{7}$$

Common knowledge of valuations may seem an unrealistic assumption. I therefore relax this assumption by assuming that valuations are private information and that the distribution of valuations is common knowledge.

Let $v_i \sim_{iid} F(v)$ with support on the interval $[0, \bar{v}]$. At the time of exerting effort, each student knows N, K, v_i and $F(v)$. The expected payoff of student i is

$$\mathbb{E}[u_i] = P_i v_i - e_i$$

where P_i equals the probability that student i passes the class, which equals the probability that S_i is one of the highest K scores.

The amount of effort student i exerts is such that the following first order condition

[4] The formal proof is given in Clark and Riis (1998) pp 286-88. The intuition is the following. Suppose there is a Nash equilibrium in pure strategies. Because there is perfect information, all students know with certainty whether they are going to pass the class. Hence, students who know they are not going to pass should exert zero effort. Because N-K students are exerting zero effort, the K remaining students should exert $e^* = \varepsilon$ with $\varepsilon > 0$ small. As a result, it is not optimal for the non active students to exert zero effort, because they can outbid the active students and get a positive utility.

[5] The K+1 students with highest valuations exert e_i from probability distribution functions $H_i(e)$ over $[v_i^l, v_{K+1}]$, with common upper support $v_i^u = v^u = v_{K+1}$, and lower support given by

$$v_i^l = \begin{cases} 0 & \text{if } i = K+1 \\ [1 - \prod_{j=1}^{K} (\frac{v_j}{v_i})] v_{K+1} & \text{if } i = 1, 2, ..., K \end{cases}$$

holds:

$$\frac{\partial \mathbb{E}[u_i]}{\partial e_i} = v_i \frac{\partial P_i}{\partial e_i} - 1 = 0$$

Athey (2004) proves the existence of pure strategy Nash equilibria in this setting. She first defines the single crossing condition which requires that, for every player i, when each of player i's opponents uses a nondecreasing pure strategy, player i's expected payoffs satisfy Milgrom and Shannon's (1994) single crossing property. She shows that the single crossing condition implies existence of a pure strategy Nash equilibrium and then she characterizes the single crossing condition based on properties of the primitives. She shows that, in the all pay auction with independent private values and (weakly) risk averse bidders, a pure strategy Nash equilibrium exists in nondecreasing strategies.

Barut el al (2002) show that the conditions of the Revenue Equivalence Theorem are satisfied in this setting. Therefore, using the Revenue Equivalence Theorem, the equilibrium expected level of effort under relative grading is

$$W^R = K \int_0^{\bar{v}} v_{K+1} dF_{K+1}(v_{K+1}) \tag{8}$$

For example, when abilities are uniformly distributed, the equilibrium expected level of effort is the following[6]

$$W^R = \frac{\bar{v}(N-K)K}{N+1} \tag{9}$$

As with relative grading, an increase in N keeping K constant has a positive effect in total effort, while an increase in K leaving N constant has an ambiguous effect. By increasing N, there are more high ability students, so v_{K+1} is larger. By increasing K, v_{K+1} decreases and K increases, so by looking at equation (8) we can see that the effect is ambiguous.

An increase in N decreases the probability of passing the class, so high ability students need to exert more effort, but there are also some students who give up and exert less effort. On the other hand, an increase in K increases the probability of passing the class, so high ability students need to exert less effort, but the increase in K also encourages some middle ability students to exert more effort.

1.3.3. Comparing the Two Regimes

I can now compare total effort and the distribution of effort under both grading schemes. This comparison is made for the case where the expected number of students who pass the class under relative grading is equal to the number of students who pass the class under absolute grading. Although with absolute grading the teacher doesn't know exactly ex ante the exact amount of students who will pass the class, I am assuming that she knows the distribution of abilities in the class, so she can set the threshold S^* to control how many students in expectation will pass the class. As argued in Dubey and Geanakoplos (2005), "By virtue of repeated meetings of the class, or similar classes held over many years, it is not unreasonable to suppose that this distribution can be fairly well estimated

[6] When F is the uniform distribution with support [0,1], $v_{K+1} \sim Beta(N-K, K+1)$, so $\mathbb{E}[v_{K+1}] = \frac{N-K}{N+1}$

by the professor and students alike".

Result 1.5 *Suppose the teacher wants $E[\#prizes] = K$. When the production function is $S_i = e_i$, the expected total level of effort under absolute grades is higher than the expected total level of effort under the relative grades.*

Proof See the Appendix. ∎

Result 1.6 *Suppose the teacher wants E[#prizes] = K and the production function is $S_i = e_i + \theta$. There is a range of values of σ_θ for which $W^A > W^R$. This range depends on $v_K - v_{K+1}$.*

Proof Follows from Result 1.5, Result 1.3 and Result 1.4. ∎

Result 1.7 *Suppose the teacher wants $E[\#prizes] = K$ and the production function is $S_i = e_i$. For students with $v_i < S^*$, the level of effort is (weakly) lower when grades are absolute. The opposite is true for students with $v_i > S^*$.*

Proof See the Appendix. ∎

Result 1.8 *Suppose the teacher wants E[#prizes] = K and the production function is $S_i = e_i + \theta$. For students with $v_i < S^*$, the level of effort is (weakly) lower with absolute grading. For students with $v_i \geq S^*$, there is a range of values of σ_θ for which the level of effort is higher with absolute grading. In particular, if $W^A \geq W^R$, then $v_i \geq S^*$ for high ability students (sufficient condition).*

Proof Same as proof of Result 1.7 ∎

To illustrate the previous results, I consider the optimal effort as a function of student type for the case of no uncertainty and 30 students, where 20 students pass the class. Effort is illustrated in Figure 1. The solid line, which represents effort under relative grades, is higher for the lower ability students compared to the dashed line, which is effort under absolute grades. The opposite is true for the high ability students. The crossing point corresponds to Kth valuation. Figure 2 graphs total effort as a function of K under both regimes, for $\theta \sim N(0, \sigma_\theta^2)$. The total level of effort under the relative standard is a function of σ_θ, as shown in the graph. The total level of effort under the absolute standard is equal to v_K when $\sigma_\theta = 0$, and for this particular combination of parameters, is decreasing in σ_θ. Total effort is equal under both regimes for $\sigma_\theta = 0.103$.

Results 1.5 and 1.7 compare relative and absolute grading under the same number of prizes, i.e, same K. This is a limitation because the optimal number of prizes could differ for the different schemes. Result 1.9 compares both regimes for their optimal K.

Result 1.9 *Let $S_i = e_i$ and let K^* maximize total effort under a relative grading system, i.e.*

$$K \in argmax_K W^R$$

Then there exists a S^* such that $W^A(S^*) > W^R(K^*)$.

Proof Follows directly from Result 1.5. ∎

Figure 1: Effort by Grading System

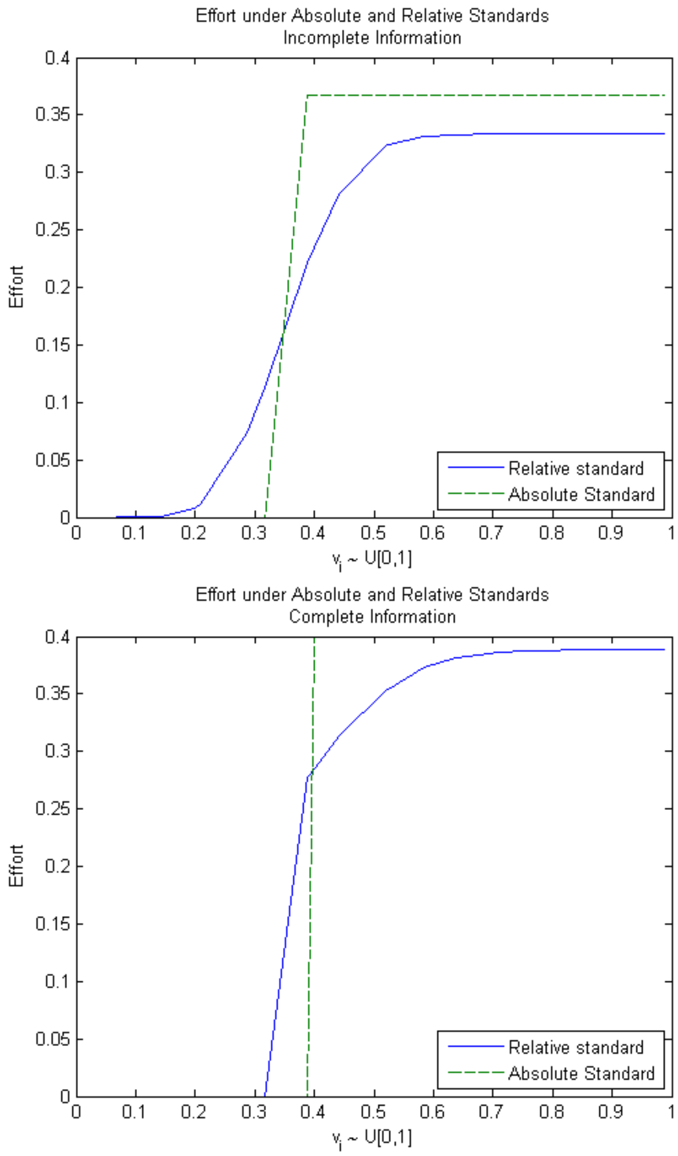
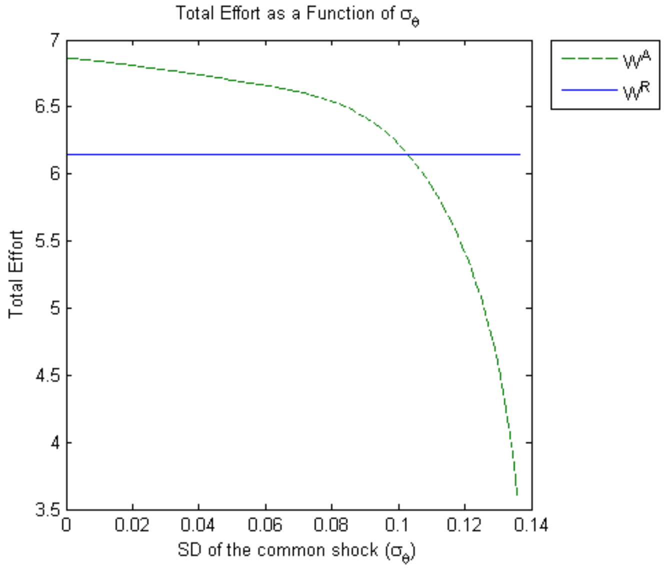


Figure 2: Effort as a Function of Uncertainty



Result 1.9 implies that the teacher can always induce higher total effort with the absolute standard than when using a relative standard.

The previous results are for a finite N, which I will assume in this paper. If N tends to infinity, the differences between the two grading systems tend to disappear when there is no uncertainty, as shown in Result 1.10.

Result 1.10 *When $S_i = e_i$ and $N \to \infty$, the effort distribution is the same under the two systems.*

Proof When $N \to \infty$, student's i problem can be written as follows

$$\max_{e_i} v_i Pr(e_i \geq \xi_p) - e_i$$

where ξ_p is the population quantile function of e_i. The expected amount of effort is the following

$$e_i = \begin{cases} 0 & \text{if } v_i < \xi_p \\ \xi_p & \text{if } v_i \geq \xi_p \end{cases} \tag{10}$$

Then $S^* = \xi_p$ and the result follows. ∎

Result 1.11 *When $S_i = e_i + \theta$ and $N \to \infty$, total expected effort is larger under the relative standard. All types of students exert lower effort under the absolute standard.*

Proof When the standard is relative, the problem the student faces is

$$\max_{e_i} v_i Pr(S_i \geq S_{(K+1)}) - e_i \quad S_i = e_i + \varepsilon_i$$

where $S_{(K+1)}$ is the K+1 order statistic of $S_1, S_2, S_3, ..., S_N$. With N sufficiently large, there is no strategic component in these games: student i knows he has no effect on $S_{(K+1)}$. Because there is no strategic component, $e_1, e_2, e_3, ..., e_N$ are independent. Moreover, if we consider a symmetric equilibrium, $e_1, e_2, e_3, ..., e_N$ are i.i.d. Therefore, using asymptotic theory of sample percentiles for the i.i.d. case, we know that as $N \to \infty$, $Var(S_{(K+1)}) \to 0$. Then students i's problem can be written as follows

$$\max_{e_i} v_i Pr(e_i \geq \xi_p) - e_i$$

where ξ_p is the population quantile function of e_i.

Because the relative standard provides insurance against the common shock, there is more variance when the standard is absolute. In that case, the teacher needs to lower S^* so that the same expected amount of students will win under both regimes. Because $S^* < \xi_p$, high ability students exert less effort. This, plus the fact that low ability students exert zero effort under absolute standards, proves the result. ∎

1.4. Empirical Strategy

My model shows that the grading system could influence both the total amount of effort in a class and the level of individual effort throughout the ability distribution, but the effort gains of moving from an absolute to a relative system cannot be uniquely signed. Which system leads to greater effort globally or on a different point of the ability distribution remains thus an empirical question.

Addressing this question in a causal way requires exploiting an experimental or quasi experimental design where one could observe both types of grading. The University of Chile provides a promising quasi experimental case. During the first semester in the Faculty of Business and Economics of the University of Chile, students have to take one course in Algebra and one course in Calculus. Before 1998, the grades in these classes were absolute grades. Beginning in 1998, the grading policy changed to relative grading and, in 2003, it changed back to absolute grading. One could thus use the grade distribution before and after 1998-2002 to identify the causal impact of relative grading, provided that there are no cohort effects. If this assumption holds, then cohorts in the period 1990-1997 are a good control group for cohorts in the 1998-2002 period.

Following the model, the score of student i is a function of effort and a shock,

$$S_i = e_i(rel, N, K, ability) + \theta_k$$

where

$$rel = \begin{cases} 0 & \text{if grades are absolute} \\ 1 & \text{if grades are relative} \end{cases}$$

The average score in a section, $\bar{S}$, is a function of average effort, $\bar{e}$, which according to the model is a function of the grading system, N, K, the distribution of ability and the variance of the shock.

$$\bar{S} = \bar{e}(rel, N, K, F(ability), \sigma_\theta)$$

The model has the following predictions for the individual scores:

1) $\frac{\partial S_i}{\partial ability} > 0$
2) For small values of σ_θ, $\frac{\partial S_i}{\partial rel} < 0$ for high ability students and $\frac{\partial S_i}{\partial rel} \geq 0$ for low ability students.

3) $\frac{\partial S_i}{\partial percentagepass} < 0$ for high ability students and $\frac{\partial S_i}{\partial percentagepass} \geq 0$ for low ability students

The model also gives predictions for the average score in a section:

4) $\frac{\partial \bar{S}}{\partial rel} < 0$ for σ_θ sufficiently low.
5) $\frac{\partial \bar{S}}{\partial N} > 0$
6) $\frac{\partial \bar{S}}{\partial K} \gtrless 0$

For the average data, the linear regression I estimate is the following:

$$\overline{Grades}_s = \delta_0 + \delta_1 \overline{ability}_s + \delta_2 \sigma_{ability} + \delta_3 rel_s + \delta_4 percentagepass_s + \varepsilon_s \tag{11}$$

where s denotes the section. Instead of the distribution of ability, I use the mean and the standard deviation, which is a good approximation if the distribution is normal. The model predicts $\delta_1 > 0$. δ_4 could be either positive or negative. A positive δ_3 would be evidence of a large σ_θ, while a negative δ_3 would be evidence that the production function without uncertainty is a good approximation.

Assuming K/N is constant, the linear regression for the individual data is

$$Grades_i = \alpha_0 + \alpha_1 ability_i + \alpha_2 rel + \alpha_3 relXability + \varepsilon_i \tag{12}$$

If σ_θ is small, the model predicts that $\alpha_1 > 0$, $\alpha_2 > 0$ and $\alpha_3 < 0$.

If K/N is not constant between the relative grading period and the absolute grading period, then α_2 and α_3 will capture the effect of the grading system and the effect of the change in the percentage of students who pass the class. If the percentage of students who pass the class is larger under relative grading, then α_2 and α_3 should be considered upper bounds of the true effect of relative grading. If the percentage of students who pass the class is larger under absolute grading, then α_2 and α_3 are lower bounds.

Suppose there are two periods with absolute grading, period abs_1 where the percentage of students who pass is lower that the percentage in the period with relative grading, and period abs_2, where the opposite is true. Then the linear regression is the following

$$Grades_i = \beta_0 + \beta_1 ability_i + \beta_2 rel + \beta_3 relXability + \beta_4 abs_2 + \beta_5 abilityXabs_2 + \varepsilon_i \tag{13}$$

The true effect of relative grading on the intercept is $\alpha_2 \in [\beta_2 - \beta_4, \beta_2]$ and the true effect of relative grading on the slope is $\alpha_3 \in [\beta_3 - \beta_5, \beta_3]$, where β_4 and β_5 are the effects (on the intercept and slope) of changing the percentage of students who pass from the levels in period abs_1 to the levels in period abs_2.

Both equation (12) and (13) assume that K and N are exogenously given.

Structural Estimation

For absolute grading, and a normal shock, my model predicts that student i's grade is given by $S_i = e_i + \theta_k$, where

$$S_i = \begin{cases} S^* + \sigma_\theta x_i + \theta_k & \text{if } v_i \Phi(x_i) - S^* - \sigma_\theta x_i \geq v_i \Phi(\frac{-S^*}{\sigma_\theta}); \\ \theta_k & \text{otherwise.} \end{cases}$$

and $x_i = \sqrt{2ln(\frac{v_i}{\sqrt{2\pi\sigma_\theta^2}})}$

To give more flexibility to the model, I allow the shock to be different if $e_i = 0$. When $e_i = 0$, let the shock be $\eta_k \sim N(0, \sigma_\eta^2)$. Then

$$S_i = \begin{cases} S^* + \sigma_\theta x_i + \theta_k & \text{if } v_i\Phi(x_i) - S^* - \sigma_\theta x_i \geq v_i\Phi(\frac{-S^*}{\sigma_\eta}); \\ \eta_k & \text{otherwise.} \end{cases} \tag{14}$$

where $x_i = \sqrt{2ln(\frac{v_i}{\sqrt{2\pi\sigma_\theta^2}})}$

The teacher chooses the standard such that

$$v_K\Phi(x_K) - S^* - \sigma x_K - v_K\Phi(\frac{-S^*}{\sigma_\nu}) = 0 \tag{15}$$

For computational reasons, instead of using equation (15) to estimate the standard, I use the following:

$$S^* = \alpha v_K \tag{16}$$

Because v_i, S_i and S^* could be measured using different scales, I will incorporate this in the model by distinguishing between real grades and nominal grades. I will define grades as real when they are measured using the same scale as v_i. S_i and S^* in equations (14) and (16) are measured in real terms. Nominal grades can be measured in any scale. Then

$$Grades_i = S_i^N = \lambda S_i \tag{17}$$

I will assume that ability can be decomposed into an entry score and an error term:

$$v_i = EntryScore_i + m_{il} \tag{18}$$

where $l \in \{1, 2, 3\}$ are three different type of students, with $v_i \sim N(\mu_v, \sigma_v^2)$ and $m_{il} \sim N(\mu_{m_l}, \sigma_{m_l}^2)$.

The structural model is given by equations (14), (16), (17) and (18). I use GMM to estimate σ_θ, σ_η, σ_v, μ_v, σ_{m_l} and μ_{m_l} for $l \in \{1, 2, 3\}$, the proportion of students of each type, π_1, π_2 and π_3, and α. I normalize $\mu_{m_3} = 0$, so there is a total of 12 parameters to estimate. The moments I use are the number of students who pass, the mean and variance of entry scores and grades, the covariance between entry scores and grades, and percentiles 5 to 95 of the grade distribution.

1.5. Data

This study employs data from the universe of students who studied in the Faculty of Business and Economics in the University of Chile in 20 consecutive years up to 2009. The Faculty of Business and Economics holds two 5-year undergraduate programs, IC and IICG. During the first year, all students have to take Algebra and Calculus classes. Students are assigned to sections of 50-70 students and each section is assigned to an instructor.

Table 1: Absolute versus Relative Grading Systems

	Absolute Grading	Relative Grading
Range	1 to 7	1 to 7
Min grade	1	average between 1 and minimum grade before relativization
Max grade	7	average between 7 and maximum grade before relativization
Passing grade	4	mean before relativization minus 1/4 of the standard deviation

The entry score is given by a standardized performance test, the Academic Aptitude Test for cohorts up to 2004, and the University Selection Test for cohorts from 2005 to 2009. These tests are college entrance examinations analogous to the SAT in the United States. Scores on both tests fluctuate between 0 and 850 points. The average of each test is set at 500 points and then the scores are adjusted to a normal distribution with a standard deviation equal to 100 points.

Grades in Chile are on a scale from 1 to 7, where 1 is the minimum grade and 7 the maximum grade. Students pass the course if their grade is greater than or equal to 4. Under the absolute grading policy, scores and grades are equivalent. Under the relative grading policy, scores are transformed to grades using the following rule. A grade of 4 is assigned to the mean score minus 1/4 of the standard deviation. 1 is given to the average between 1 and the minimum score and 7 is assigned to the average between 7 and the maximum score. All other scores are relativized using the two lines between these three points.

For cohorts until 1997, grades were absolute. In 1998, the grading policy changed and classes were graded on a curve, until 2003, when they changed back to absolute grading.

The University of Chile provided data for cohorts starting in 1980 and ending in 2009. I only use cohorts starting in 1990 because many earlier entry scores are missing. In total, I have 8801 students. From this, I keep students who completed Calculus and Algebra during their first year and whose grades are not missing. I also drop students who enter the University by special admission, since they don't have comparable entry scores. In total, I have 6006 students and 12004 observations. 34.89% of the observations are for the period 1990-1997, 26.72% for the period 1998-2002 and 38.39% for the last period. The 6006 students are divided into 416 sections, 201 for Algebra and 215 for Calculus. 89 sections use relative grading, while the other 327 have an absolute grading policy.

In my data set I don't observe effort directly, but I can observe it indirectly through grades. I thus use the grade distribution in sections before and after 1998-2002 to identify the causal impact of the new grading system. The assumption that cohorts before 1998 are a good control group for cohorts between 1998-2002 is a good assumption provided that there is no difference in selection into the Business and Economics Faculty. On one hand, the change in the grading policy is unlikely to produce a different selection, since the change in regime was only for the math classes, which are only 4 out of nearly 50 courses. On the other hand, because we are looking at a long period of time, there could be other factors that may produce different selection. This is a limitation of the data.

Table 2: Grading System, Entry Score, Percentage Male and Passing Rate by Period

Period	Grading	Entry Score	% Male	% Pass per Section
1990-1997	Absolute	688.72	0.576	0.533
		(23.05)	(0.494)	(0.165)
		4188	*4188*	*4188*
1998-2002	Relative	689.74	0.590	0.707
		(18.19)	(0.492)	(0.075)
		3208	*3208*	*3208*
2003-2009	Absolute	701.15	0.606	0.803
		(24.54)	(0.489)	(0.128)
		4523	*4608*	*4608*
Total		693.71	0.592	0.683
		(23.20)	(0.492)	(0.176)
		11919	*12004*	*12004*

Notes: Standard deviation are presented in parentheses and number of observations in italics

1.6. Results

I start by using aggregate data (at the section level) to estimate equation (11). Because in practice, the percentage of students who pass the class is likely to be endogenous at the section level, I use the percentage of students who pass the class at a year level[7]. Results are presented in column (2) of Table 3. As expected, a higher entry score translates in higher grades. The coefficient of relative grading is positive but statistically insignificant, so there is no clear evidence that the grading system has an impact on the average level of effort in a class. In column (3), I add controls for the percentage of male students in the classroom, which doesn't change the relative grading coefficient.

There is no empirical evidence that the grading system affects the average grade in a section, but there could still be an effect in the distribution of grades. To test this, I use individual data.

Because the percentage of students who pass differs between the three different periods, I estimate linear equation (13). Results are presented in Column (1) of Table 4. There is evidence that the effect of relative grading on the intercept (α_2) is positive and the effect of relative grading on the slope (α_3) is negative, in that $\beta_2 - \beta_4 > 0$ and $\beta_3 - \beta_5 < 0$ and the F-test indicates that both are statistically different from zero.

To control for differences in observable characteristics of students between the three periods (percentage of male students per section and differences in the entry score), I follow DiNardo, Fortin, and Lemieux (1996) and use propensity score reweighting. That is, I build the counterfactual distribution of grades for periods 1998-2002 and 2003-2009 as *if the percentage of males and the distribution of the entry score was the one prevailing in period 1990-1997*, and compare these counterfactual distributions to the distribution of the grades for period 1990-1997.

The results of estimating equation (13) with DFL reweighting are shown in column (2) of Table 4. After reweighting, the difference between the β_2 and β_4 is positive and the difference between β_3 and β_5 is negative but neither is statistically different from zero.

[7] This is likely to be set by the administration and not by the instructor

Table 3: Effect of Relative Grading on Students' Grades: Aggregate Data

VARIABLES	(1) Grade	(2) Grade	(3) Grade	(4) Algebra	(5) Calculus
Mean entry score	0.0110***	0.00693***	0.00692***	0.00641***	0.00787***
	(0.00154)	(0.00101)	(0.00105)	(0.00152)	(0.00139)
Sd entry score	0.00908**	0.00314	0.00315	0.00736*	-0.000485
	(0.00427)	(0.00277)	(0.00278)	(0.00435)	(0.00347)
% Male			0.00656	-0.07	0.0649
			(0.143)	(0.220)	(0.180)
relative	0.0942	0.0644	0.0644	0.0406	0.0976*
	(0.06450)	(0.04170)	(0.04170)	(0.06220)	(0.05460)
mean % pass		2.483***	2.483***	2.419***	2.290***
		(0.118)	(0.118)	(0.170)	(0.172)
Constant	-3.646***	-2.469***	-2.465***	-2.030**	-3.029***
	(1.052)	(0.681)	(0.687)	(1.005)	(0.906)
Observations	319	319	319	159	160
R-squared	0.187	0.663	0.663	0.665	0.642

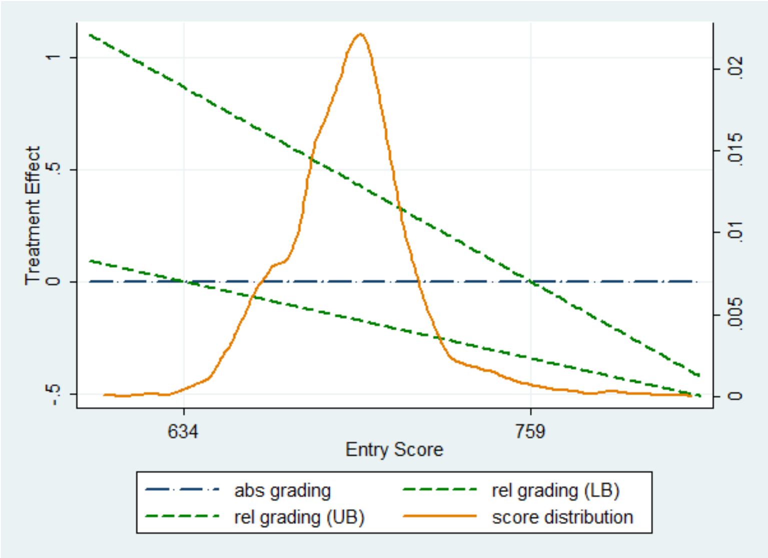
Notes: Robust standard errors are presented in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Effect of Relative Grading on Students' Grades: Individual Data

VARIABLES	(1) Grade	(2) Grade DFL rw	(3) Algebra	(4) Algebra DFL rw	(5) Calculus	(6) Calculus DFL rw
Relative	5.945***	5.281***	5.122**	4.460**	6.789***	6.121***
	(1.478)	(1.454)	(2.166)	(2.145)	(1.919)	(1.869)
Entry score	0.0157***	0.0157***	0.0149***	0.0149***	0.0166***	0.0166***
	(0.00172)	(0.00172)	(0.00260)	(0.00260)	(0.00208)	(0.00208)
relative*score	-0.00795***	-0.00696***	-0.00688**	-0.00589*	-0.00905***	-0.00805***
	(0.00216)	(0.00212)	(0.00316)	(0.00313)	(0.00280)	(0.00272)
$D_{2003-2009}$	3.367**	3.537**	2.122	2.474	4.683***	4.621***
	(1.310)	(1.406)	(1.932)	(2.054)	(1.564)	(1.701)
$(D_{2003-2009}) * score$	-0.00397**	-0.00421**	-0.00208	-0.00258	-0.00596***	-0.00587**
	(0.00190)	(0.00204)	(0.00281)	(0.00299)	(0.00227)	(0.00247)
Constant	-7.159***	-7.159***	-6.488***	-6.488***	-7.850***	-7.850***
	(1.179)	(1.179)	(1.779)	(1.779)	(1.427)	(1.427)
Observations	11,598	11,598	5,833	5,833	5,765	5,765
R-squared	0.155	0.132	0.173	0.142	0.144	0.13

Notes: Standard errors, clustered at the section level, are presented in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure 3: Effect of Relative Grading



From the point estimates, the lower and upper bound for the effects on the intercept are 1.704 and 5.281 respectively, and the lower and upper bound for the effects on the slope are -0.0027 and -0.00696.

The results of column (2) in Table 4 are graphically presented in Figure 3. The X-axis represents the prediction of the grade given the student's entry score under absolute grading and the percentage of students who passed during the period 1990-1997. The two other lines represent the upper and lower bound of changing the grading system from absolute grading to relative grading. The lower bound shows that all students with entry scores below 634 exert more effort after the change, while students with entry score above 634 exert less effort with relative grading. The upper bound shows that all students with entry scores below 759 exert more effort after the change. Thus, taking the two sets of results together, students with entry score under 634 are getting higher grades with relative grading and students with entry score above 759 are getting lower grades under relative grading. This corresponds to the 1st and 99th percentile of the ability distribution. For the remaining students, we need a precise estimate of the effect of relative grading to asses whether they are better or worst off after the change.

To get a more precise estimate, I use the following assumption:

Assumption 1.12 *The effect of changing the percentage of students who pass by 1 percentage point is constant.*

Assumption 1.12 means that the effect on grades of increasing the percentage of students who pass the class by 1 percentage point is the same no matter the initial number of students who pass the class. This assumption seems unrealistic for very low and high initial levels, but is reasonable for an intermediate range of initial levels.

Since the difference in the percentage of students who pass the class between the first and second period is 64% of the difference between the first and third period, then the

Table 5: Effect of Relative Grading on Students' Grades: Upper and Lower Bounds

		Coef.	Std. Err.
Upper Bound	β_2	5.280617	(1.453591)
	β_3	-0.00696	(0.002116)
Lower Bound	$\beta_2 - \beta_4$	1.743775	(1.145077)
	$\beta_3 - \beta_5$	-0.00275	(0.001655)
Assumption 1.12	$\beta_2 - 0.64\beta_4$	3.017038	(1.069666)
	$\beta_3 - 0.64\beta_5$	-0.00427	(0.00155)

$Grade_i = \beta_0 + \beta_1 ability_i + \beta_2 rel + \beta_3 rel * ability + \beta_4 abs_2 + \beta_5 ability * abs_2 + \varepsilon_i$

effect of relative grading under Assumption 1.12 is $\beta_2 - 0.64\beta_4$ and $\beta_3 - 0.64\beta_5$. In that case, all students with entry score above 707 points are exerting less effort after the change. This corresponds to the 25th percentile of the entry score distribution. Results for the lower and upper bound and estimates under Assumption 1.12 are presented in Table 5 and in Figure 5.

Figure 4 shows the grade distribution for the three periods, with and without reweighting. The mean grade is similar for the three periods, which is consistent with the results in Table 3. In the 1990-1997 period, the number of students with low grades is higher compared with the other periods, and the number of students with high grades is lower. When I use propensity score reweighting, the difference between period 1998-2002 and period 2003-2009 seems negligible. Because the percentage of students who pass is almost 10 percentage points higher in the 2003-2009 period than in the 1998-2002 period, these results suggest that the effect of relative grading is as strong as the effect of increasing the percentage of students who pass the class by 10 percentage points.

1.6.1. Structural Estimation Results

The results of the structural estimation for effort under absolute grading are shown in Table 6. To see the fit of the model, Table 7 shows the difference between the actual and simulated moments. The variance of grades is greater in the simulations than in the actual data, while the opposite is true for the mean grade. The predicted grade for percentiles 5 to 25 is lower in the model than in the actual data, but the fit is good for percentiles 30 to 90. Overall, the model has a good fit except for the lower tail of the grade distribution.

The model predicts that, when students know that they cannot pass the class, they should exert no effort. However, the data shows that this doesn't always hold, as there are students that fail the class with very low grades. This results in a poor fit for the lower tail of the grade distribution. To improve the fit, a dynamic model is needed, where students can adjust their level of effort between tests. These extensions of the model are left for future research.

1.6.2. Policy Simulations

Becker and Rosen (1992) argue that stratifying students into groups in which each has a chance for success may be preferred to a single standard. I use my structural model to

Figure 4: Grade Kernel Densities

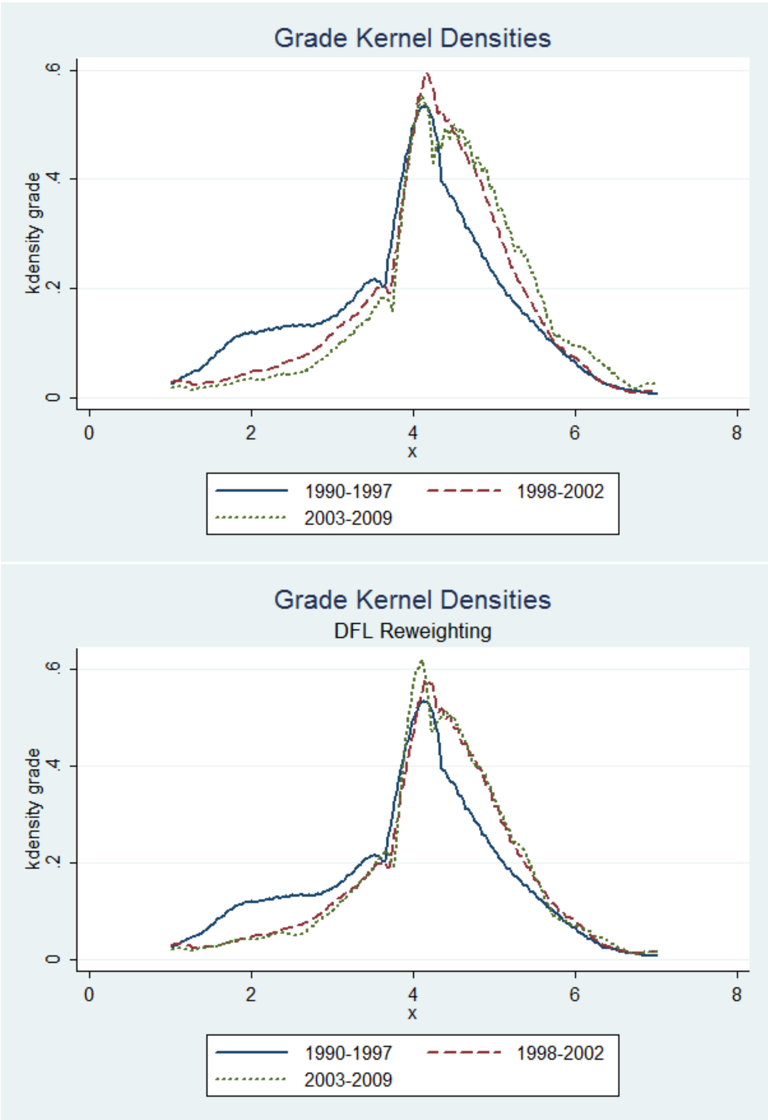


Figure 5: Effect of Relative Grading

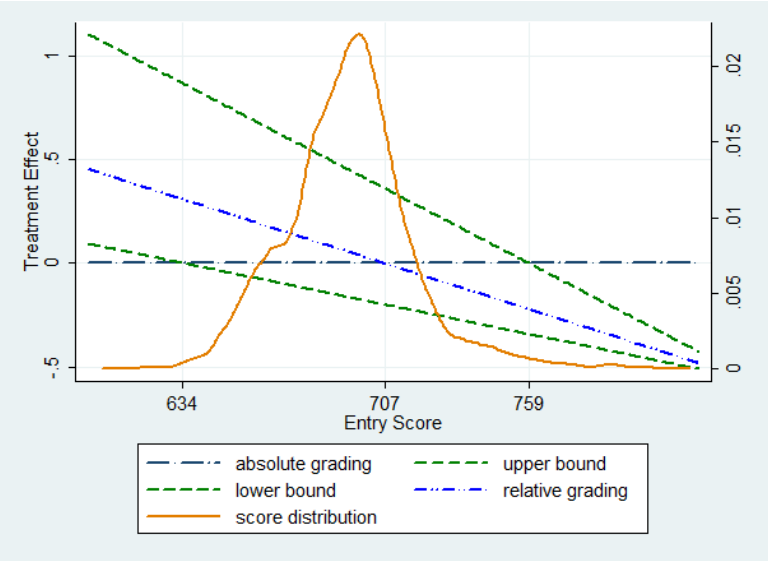


Table 6: Structural Parameters

	Parameter	Standard Dev
α	0.83	(0.041)
σ_θ	18.20	(3.449)
σ_η	522.39	(80.171)
σ_v	24.44	(0.191)
μ_v	694.24	(1.940)
σ_{m_1}	84.82	(7.379)
σ_{m_2}	$1.34X10^{25}$	$(2.40X10^{48})$
σ_{m_3}	64.49	(20.134)
μ_{m_1}	-36.79	(17.523)
μ_{m_2}	222.80	(51.497)
π_1	0.33	(0.000)
π_2	0.33	(0.000)

Table 7: Model Fit: Actual versus Simulated Moments

Moments	Actual	Simulated	Difference	Standard Dev
No of passes	5904	5912	8.00	(5.260)
Cov(Si,vi)	9.73	9.82	0.09	(0.458)
Var(vi)	605.39	605.57	0.19	(0.717)
Var(Si)	1.34	2.48	1.14	(0.313)
Mean(vi)	695.12	694.61	-0.51	(1.935)
Mean(Si)	4.09	3.69	-0.40	(0.148)
p05	1.9	1.00	-0.90	(0.000)
p10	2.4	1.00	-1.40	(0.000)
p15	2.8	1.00	-1.80	(0.010)
p20	3.2	1.32	-1.88	(0.180)
p25	3.5	2.47	-1.03	(0.203)
p30	3.7	3.53	-0.17	(0.251)
p35	4.0	4.08	0.08	(0.099)
p40	4.0	4.16	0.16	(0.099)
p45	4.1	4.20	0.10	(0.101)
p50	4.2	4.25	0.05	(0.103)
p55	4.3	4.29	-0.01	(0.107)
p60	4.4	4.34	-0.06	(0.115)
p65	4.5	4.79	0.29	(0.268)
p70	4.7	4.90	0.20	(0.307)
p75	4.8	4.94	0.14	(0.319)
p80	5.0	4.98	-0.02	(0.328)
p85	5.2	5.02	-0.18	(0.336)
p90	5.5	5.06	-0.44	(0.344)
p95	5.9	5.12	-0.78	(0.357)

explore the effects of tracking on student effort.

Specifically, I use the results from the structural estimation to study what would have happened to grades in Algebra in year 2009 if students had been tracked. In this year, there were only eight sections. Table 8 shows the mean entry score, percentage of males and percentage of students who pass by section. There are no significant differences in the mean entry score between sections, which is consistent with no tracking.

If students were assigned to sections based on their ability, the model predicts that effort would increase. The problem is that ability is unobserved. In the structural model, I assumed that ability could be decomposed into the sum of an entry score and some unobserved component. Therefore, I will study the effects of tracking by assigning students to sections based on their entry score.

Table 9 shows descriptive statistics of the actual grade distribution and the results from the simulations with and without tracking. The assumption is that the expected number of students who pass by section stays the same with and without tracking. As discussed before, the simulated mean grade is lower than the actual grade and the simulated standard deviation is higher. The model does not fit too well at the lower tail of the grade distribution, but otherwise provides a satisfactory fit. When students are tracked, the mean grade increases by 0.07 points, and the standard deviation decreases by 0.17 points. Tracking also increases the simulated grade for all percentiles, except for the median and percentile 95. Overall, the effect of tracking on grades is positive, even if ability is unobserved.

Table 8: Descriptive Statistics by Section: No Tracking

Section	Entry score	male	pass
1	720.12	0.57	0.96
	(16.88)	(0.50)	(0.19)
2	712.44	0.62	0.86
	(19.71)	(0.49)	(0.35)
3	716.76	0.71	0.83
	(32.21)	(0.46)	(0.38)
4	727.14	0.56	0.94
	(17.49)	(0.50)	(0.24)
5	726.04	0.62	0.81
	(22.76)	(0.50)	(0.40)
6	722.26	0.71	1.00
	(26.37)	(0.46)	(0.00)
7	719.55	0.57	0.89
	(16.19)	(0.50)	(0.31)
8	716.68	0.74	0.93
	(18.16)	(0.45)	(0.27)

Table 9: Effect of Tracking on Grades

	Actual		Simulated	
Moments		No tracking	Tracking	Difference
Mean	4.74	4.30	4.37	0.07
Std Dev	0.88	1.29	1.12	-0.17
p1	1.90	1.00	1.00	0.00
p5	3.00	1.00	1.81	0.81
p10	4.00	1.78	2.90	1.12
p25	4.35	4.20	4.23	0.02
p50	4.80	4.78	4.38	-0.40
p75	5.20	4.98	5.05	0.07
p90	5.70	5.08	5.36	0.28
p95	6.00	6.49	5.36	-1.13
p99	6.50	6.49	6.56	0.07

1.7. Conclusions

Both my model and the empirical evidence for the Chilean case show that the grading system affects the decision of students about how much effort to exert. Although I didn't find empirical evidence that the grading system has a significant effect on the aggregate level of effort, I showed that the distribution of effort is affected.

I find that low ability students, with ability measured as the entry score to the University, exert higher effort when the grading system is relative. For example, a student with an entry score of 690 would increase his grade by 0.07 points out of 7 when the grading system changes from absolute to relative, according to the linear estimation.

On the other hand, high ability students lower their effort when the grading system changes from absolute to relative. For example, a student with an entry score of 720 would decrease his grade by 0.057 points out of 7 with the change in the grading system.

The welfare implications of the grading system will depend on the objective function of the teacher. When students care only about passing the class, the teacher has to decide whether to focus on low or high ability students. If the aim of the teacher is to provide incentives for lower ability students to exert more effort, the relative grading system seems to be the preferred one. If the objective of the teacher is to identify and focus her efforts on the high ability students, then grades should be absolute.

Tracking is a promising solution to provide incentives to both groups of students, even when true ability is unobserved and tracking is done based on a proxy for ability. In the case of the University of Chile, this can be easily done by assigning students to different sections based on their entry score. My simulations showed that, for year 2009, tracking could have raised the mean grade by 0.05 standard deviations.

This paper does not evaluate other possible grading systems, so I cannot say which is the optimal grading system. I will leave this question for future research.

2. Grading Standards and Student Achievement

2.1. Introduction

The recent empirical literature on teacher quality seems to agree that high quality teachers are the most important assets of schools (Hanushek and Rivkin, 2010). However, it hasn't been possible to identify any measurable characteristic of teachers that is reliably related to student outcomes. With the exception of the first years of experience, the evidence on the effect of teacher certification, master's degree and training is mixed, with studies such as Harris and Sass (2006), Hanushek and Rivkin (2004, 2006), Boyd et all (2006) and Kane, Rockoff and Staiger (2008) finding no effect.

While most studies have focused on identifying characteristics of teachers that predict teacher quality, less attention has been given to identifying effective teacher practices. Lavy (2011), Tyler et al. (2010) and Kane et al. (2011) are a few exceptions. Lavy (2011) studies the effectiveness of five measures of teaching practices, where the evaluation is done by students: instilment of knowledge and enhancing comprehension; instilment of applicative, analytical and critical skills; instilment of the capacity for individual study; transparency and fairness and; feedback and individual treatment of students. He finds that the first practice has a very strong and positive effect on test scores, particularly among girls and pupils of low socioeconomic background, while the second practice has a very large positive payoff, especially among pupils from educated families. He finds no effect of the third practice. Other actions that lead to achievement gains are transparency in the evaluation of pupils, proper and timely feedback to students, and fairness. Kane et al. (2011) studies the Cincinnati Teacher Evaluation System, which is a peer review evaluation that focuses on four domains: planning and preparing for student learning, creating an environment for student learning, teaching for student learning, and professionalism. They find that the teacher's overall score is important in increasing student achievement.

Another action that teachers undertake is grading. Teachers usually decide the grading standards, i.e. how difficult it is to get a certain grade. A small theoretical literature (Kang, 1985, Betts, 1995, Becker and Rosen, 1992) has focused on how educational standards have an impact on students' incentives. Its main finding is that higher standards may help some students while hurting others, because as standards move beyond their reach, some students give up (Betts and Grogger, 2003). Thus, through manipulating standards, teachers can provide incentives to some students to exert effort.

One difficulty when studying the effects of grading standards is having a good measure of standards. Babcock (2009) measures standards at a class level as the average expected grade. He uses course evaluations from UC San Diego to see how average study time is affected by grading standards. He finds that average study time would be about 50% lower in a class in which the average expected grade was an A versus a class in which the average expected grade was a C.

Betts (1995) proposes to measure standards at a school level as the school fixed effect in a regression of standardized test scores on GPA and school dummies. In that case, a school with a large estimated intercept has high standards, as it assigns low letter grades relative to the student's standardized test. He finds that grading standards have a positive effect on scores, and that the effect is larger for students who obtain higher letter grades. Betts and Grogger (2003) use a very similar definition of grading standards and study the distributional effects of increasing grading standards. They find that high standards increase the test scores for all students but that the increase is greatest for the top of the test distribution. They also find no effect on educational attainment and a negative effect on high school graduation for blacks and Hispanics.

Figlio and Lucas (2004) study the effects of teacher-level grading standards on student achievement. They also find that the students who benefit most from higher standards are the initially high achieving students. In addition, they find that low achieving students benefit more from high standards when their classmates are high achieving and that high achieving students benefit more from higher standards when their classmates are low achieving.

The methodology proposed by Betts (1995) has three main problems. First, there is an endogeneity problem because schools can choose standards based on their students' performance. Second, there is a selection problem, because parents can choose schools with certain standards based on the students' ability. These two problems are dealt with in Betts and Grogger (2003) by controlling for students' past scores and the mean past score in each class.

A third problem that hasn't been discussed in the previous studies is the mechanical relationship between standards and test scores: standards are constructed as the difference between mean test scores and grades. Because the student's test score is used to compute the mean score, there is a mechanical relation that could bias the results.

This study presents new evidence on the effect of grading standards on scores. The study contributes to the literature in two ways. First, this study adopts an identification strategy that exploits the fact that we observe each student three times: we observe standardized test scores and grades for language, math and sciences. This implies that the effects of the grading standard can be identified in models that control for the influence of unobserved student traits, which may have biased the conventional cross-sectional evaluations. Second, I use instrumental variables to address the mechanical relationship between the estimated standard and test scores.

The remainder of this discussion is structured as follows. Section 2.2 discusses the data. Section 2.3 develops the model and discusses the econometric framework. Section 2.4 presents the results and Section 2.6 concludes.

2.2. Data

To estimate the grading standards, data on standardized test scores as well as grades are needed. The data set on standardized test scores that I use is the SIMCE. The SIMCE test, created in 1998, is the main instrument with which to measure the quality of education in Chile. In 1999, the test was modified and standardized in order to follow up on school performance. The new test has an open-ended scale that measures student abilities (cognitive skills). It uses the Item Response Theory (IRT) which is applied in most international tests of academic achievement (such as TIMSS or PISA) and links a student's score to his abilities. In 1999, it was also decided that the mean for that year would be 250, with a standard deviation of 50. The parameters of all subsequent evaluations have been calculated based on the mean and variance of 1999.

SIMCE provides census information on standardized educational performance tests,

Table 10: Descriptive Statistics

Variable	Description	Obs	Mean	Std. Dev.	Min	Max
score	student's test score in 2009 (8th grade)	350279	264.53	50.77	98.05	402.74
grade	student's GPA in 2009 (8th grade)	350279	52.88	7.88	17	70
male	indicator for male student	350279	0.48	0.50	0	1
fathersch	student's father years of schooling	350279	11.56	3.73	0	20
mothersch	student's mother years of schooling	350279	11.48	3.53	0	20
income	13 categories for student's household income	350279	4.32	3.21	1	13
past score	student's test score in 4th grade	350279	265.43	50.28	92.96	397.18
past mean	student's classroom mean test score in 4th grade	350279	264.67	26.35	124.21	361.74
municipal	indicator for municipal school	350279	0.39	0.49	0	1
voucher	indicator for voucher school	350279	0.53	0.50	0	1
private	indicator for private school	350279	0.08	0.27	0	1
math	indicator for math	350279	0.34	0.47	0	1
lang	indicator for language	350279	0.33	0.47	0	1
science	indicator for natural sciences	350279	0.33	0.47	0	1
curr	mean curriculum coverage by subject (1: not at all, 4: completely covered)	350279	2.88	0.48	1	4

characteristics of the household and characteristics of the school and teachers. In 2005, SIMCE evaluated 259,852 students registered in 4th grade, and 7,540 schools (MINEDUC 2006). This corresponds to 96% of students and 90% of schools. In 2009, SIMCE evaluated 239,745 students registered in 8th grade, and 5,814 schools (MINEDUC 2010). This is a coverage of 92.5% of students and 98.2% of schools. Eighth graders take four tests: Mathematics (Math), Language and Communication (Language), Social Sciences and Natural Sciences (Science). Fourth graders take only three tests: Mathematics, Language and Communication and Sciences. The scores for these three tests in both years, the socioeconomic characteristics of the students' households, and grades in 8th grade are available for 127,653 students.

Grades in Chile are on a scale from 1 to 7 (with one decimal point), where 1 is the minimum grade and 7 the maximum grade. Students pass the course if their grade is greater than or equal to 4. For purposes of this study, I will multiply grades by 10. Figure 6 shows the grade distribution for the sample used in this study.

Descriptive statistics of the variables used in this study are presented in Table 10.

2.3. Theoretical Framework

2.3.1. Effect of Grading Standards on Student Effort

Student i in school j has utility

$$V_{ij}(e_{ij}) = U(G_{ij}) - c(e_{ij}) \quad s.t. \quad G_{ij} = \tau_j + e_{ij} + u_{ij} \tag{19}$$

$$S_{ij} = \lambda e_{ij} + v_{ij} \tag{20}$$

where G_{ij} are students' grades, S_{ij} is the score of a standardized test, e_{ij} is effort, λ is a scaling parameter (S_{ij} and G_{ij} can be measured in different scales), $U(G)$ and $c(e)$ are twice differentiable function with $c'(e) > 0$, $c''(e) \geq 0$ and $U'(G) > 0$. Let's assume that $U(G)$ is S-shaped, with $U''(G) < 0$ for high levels of G. Let τ_j be a measure of how easy

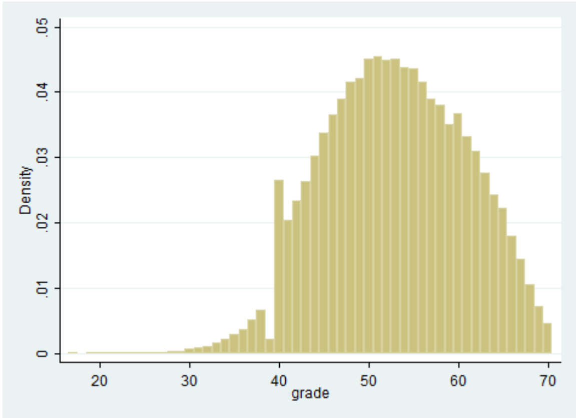
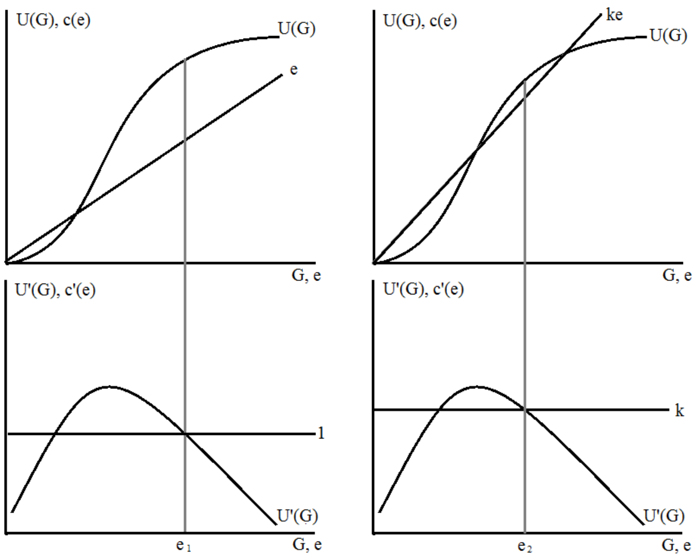


Figure 6: Grade Distribution

Figure 7: Optimal Effort



the grading in a school is. If $\tau_2 > \tau_1$, then school 2 has lower grading standards than school 1, in that students exerting the same level of effort get higher grades. Students maximize their utility by choosing effort such that

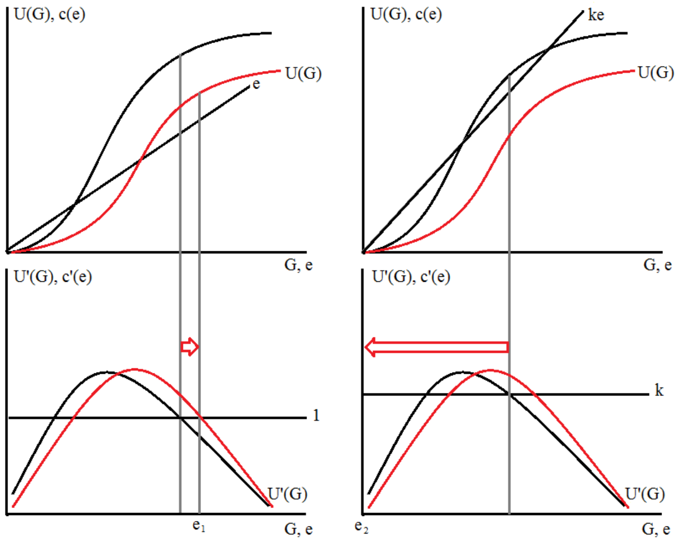
$$U'(G_{ij}) - c'(e_{ij}) = 0 \quad and \quad e_{ij} > 0, \quad or \tag{21}$$

$$U''(G_{ij}) - c''(e_{ij}) < 0 \quad and \quad e_{ij} = 0 \tag{22}$$

To see what happens to effort when τ increases, let's take the case of two students, 1 and 2, with cost functions $c_1(e) = e$ and $c_2(e) = ke$ with $k > 1$. Assume the school's grade inflation is initially τ. The effort chosen by students 1 and 2 is shown in Figure 7. Student 2 has a higher cost of exerting effort, and the function $U(G)$ is the same for both students, so $e_1 > e_2$.

Now assume the school increases the grading standard, so that $\tau' < \tau$. The optimal effort of each student is represented in Figure 8. Student 1 increases his effort, but student

Figure 8: Effect of an Increase in Standards



2 decreases his effort to zero, since he is better off giving up. Thus, the effect of increasing grading standards on effort depends on the cost of exerting effort. If we interpret a low cost of exerting effort as higher ability, then higher standards increase the level of effort for high ability students but encourage low ability students to give up and thus decrease their level of effort.

More formally, to see the effect of grade inflation on effort conditional on $e > 0$, we can differentiate the FOC given an interior solution. Then $\frac{de}{d\tau} = \frac{U''(G)}{c''(e)-U''(G)}$. Because e is a maximum, the denominator must be positive. The numerator can be either positive or negative. Because $U(G)$ is S-shaped, we know that for high initial levels of effort, we are in the concave portion of the curve and so the effect is negative.

Notice that $\frac{dS_i}{d\tau} = \frac{dS_i}{de_i}\frac{de_i}{d\tau} = \lambda\frac{de_i}{d\tau}$. In other words, because higher standards make students exert more effort, higher standards will also have a positive impact on standardized test scores. To study the effect of grading standards on test scores, we solve for effort in equation (1) and replace in equation (2):

$$S_i = \lambda G_i + \beta + \epsilon_i \tag{23}$$

where $\beta = -\lambda\tau$ is the school standard and $\epsilon_i = v_i - \lambda u_i$.

2.3.2. Estimating the Effect

Standards can be estimated at the school level, school-subject level and teacher level. Based on equation (23), I construct the standards at the school level ($Standard_s$) by estimating the following equation

$$S_{ijs} = \beta_s + \lambda GPA_{ijs} + \epsilon_{ijs} \tag{24}$$

where i indexes the student, s indexes the school, j indexes the subject, GPA_{ijs} is the student's average grade and β_s is a school fixed effect. Thus, $Standard(1)_s = \hat{\beta}_s$. If all schools graded in the same way, then $\beta_s = \beta$. If schools have different grading standards, then a school with higher β_s has higher grading standards. For example, if $\beta_1 > \beta_2$, then students in school 1 receive higher test scores than students in school 2 with exactly the same GPA. Thus, school 1 is more stringent in grading students.

I use the same procedure to estimate the standard at a school subject level, $Standard_{js}$

$$S_{ijs} = \beta_{js} + \lambda GPA_{ijs} + \epsilon_{ijs} \tag{25}$$

where β_{js} is a school subject fixed effect. Then, $Standard(2)_{js} = \hat{\beta}_{js}$

To estimate the effect of grading standards on standardized tests, I use the following specification:

$$S_{ijs} = \alpha + \gamma Standard_s + \delta_j + \varphi X_i + \psi Z_s + \eta_{ijs} \tag{26}$$

where $Standard_s$ is the grading standard at a school level, δ_j is a subject fixed-effect, X_i is a vector of student characteristics, Z_s is a vector of school characteristics and η_{ijs} is a disturbance term.

The vector X_i includes the student's gender, the student's 4th grade test score, parental income, and parental education. The term Z_i includes an indicator variables for urban schools, municipal schools, voucher schools and private schools. It also includes the average 4th grade test score for the school.

One of the concerns of estimating equation (26) is the possibility that the school's choice of grading standards may depend on students' initial skill levels (Betts and Grogger, 2003). In addition, parents may choose a school with a certain level of grading standards to maximize their child's effort. Controlling for the student's 4th grade test score and the average 4th grade test score is an attempt to deal with these issues. In addition, I exploit the fact that I observe each student three times and estimate the following equations:

$$S_{ijs} = \alpha + \gamma Standard_{js} + \delta_j + \xi_s + \upsilon_{ijs} \tag{27}$$

$$S_{ijs} = \alpha_i + \gamma Standard_{js} + \delta_j + \nu_{ijs} \tag{28}$$

where $Standard_{js}$ is the grading standard at a school-subject level, α_i is an individual fixed effect and ξ_s is a school fixed effect.

It would also be desirable to estimate the standard at the teacher level. However, the data set used in this study doesn't have a teacher identifier, so we cannot tell if the same teacher is teaching all the sections in one subject. If this were the case, then it would be better to estimate the standard at the school-subject level. Notice that $Standard_{js}$ would correspond to the grading standard at the teacher level if there is only one class per grade, which is the case for 64% of schools.

Finally, because of the way the variable *Standard* is constructed, there is a mechanical relationship between *Standard* and S

$$S_{ijs} = \alpha + \gamma(\bar{S}_{ijs} - \hat{\lambda} G\bar{P}A_{ijs}) + \delta_j + \xi_s + \upsilon_{ijs} \tag{29}$$

I use the past standard (in t-2) as an instrument for the standard in the current year. The past standard is positively correlated with the standard in the current year, and is

Table 11: Estimated Standards

Variable		Obs	Mean	Std. Dev.
Estimated Standards				
Standard(1)	School level	350279	3.79	27.29
Standard(2)	Subject level	350279	5.90	29.75
Estimated Standards by Subject				
Standard(2)	Math	350279	4.22	19.01
	Language	350279	-0.09	15.01
	Sciences	350279	1.78	17.66
Estimated Standard by Percentile				
Standard(1)	1-33%	350279	-8.52	13.58
	33-66%	350279	0.56	4.61
	66-100%	350279	11.75	18.52
Standard(2)	1-33%	350279	-8.66	14.39
	33-66%	350279	1.42	5.08
	66-100%	350279	13.15	20.91

Notes: Standard(1) refers to the grading standard measured at the school level, while Standard(2) refers to the grading standard measured at the school-subject level.

Table 12: Correlation of Estimated Standards

	Standard(1)	Standard(2)
Standard(1)	1	
Standard(2)	0.9286	1

Notes: Standard(1) refers to the grading standard measured at the school level, while Standard(2) refers to the grading standard measured at the school-subject level.

not mechanically related to S_{ijs}, since the past standard is constructed using a different cohort.

2.4. Results

2.4.1. Estimating the Standards

My identification strategy requires sufficient variation in grading standards between schools and between subjects in the same school. The results from the estimation of equations (24) and (25) suggest that this is the case. The F-test for there being no difference in grading standards between schools and between subjects is rejected at all conventional significance levels.

Table 11 shows descriptive statistics for the estimated standards. The correlations between the two estimated standards are presented in Table 12. The correlation is positive and very high (more than 90%).

Table 13 shows how characteristics of the school are correlated with standards. Rural and municipal schools grade less stringently than urban and voucher schools (0.76 and 0.90

Table 13: Standards and School Characteristics

Standards by Type of School			
Mean		**Differences**	
Urban	5.845	Urban vs Rural	20.741
Rural	-14.895		(0.956)
Municipal	-13.734	Mun vs Voucher	-24.434
Private	44.722		(1.164)
Voucher	10.699	Priv vs Voucher	34.023
			(1.002)

Notes: Standard errors in parentheses

Table 14: Estimated effects of grading standards on test scores

	OLS Regressions				IV Regressions			
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard(1)	0.709***				0.655***			
	(0.0103)				(0.0190)			
Standard(2)		0.618***	0.504***	0.503***		0.629***	0.576***	0.581***
		(0.0077)	(0.0081)	(0.0101)		(0.0161)	(0.0203)	(0.0196)
Observations	350,279	350,279	350,279	350,279	350,279	350,279	350,279	350,279
R-squared	0.569	0.576	0.338	0.84	0.569	0.576	0.037	0.076
School FE			yes				yes	
Student FE				yes				yes

Notes: Standard errors clustered at the school level are in parentheses. All models include subject fixed effects. (1), (2), (5) and (6) include controls for past test score and the past mean score. (1)-(3) and (5)-(7) include controls for mother and father's years of schooling, household income, and an indicator for male student. (1), (2), (5) and (7) include indicators for urban schools, private schools and municipal schools. Standard(1) refers to grading standard measured at the school level, while Standard(2) refers to the grading standard measured at the school-subject level.

standard deviations), while private schools have higher standards than voucher schools (1.246 standard deviations).

2.4.2. Effect on Test Scores

Table 14 presents the main results for the effect of grading standards on test scores. All standard errors are clustered at the school level. In columns (3) and (7), I include school fixed effects, and in columns (4) and (8) I include student fixed effects. I use instrumental variables for columns (5)-(8).

The results show that lower standards decrease the average test score by a statistically significant amount. Although the coefficients are positive and statistically significant in all specifications, we see that, when we control for school and student unobserved characteristics, the estimated effect of grading standards on scores decreases. This may have different explanations. On one hand, this suggests that controlling for the students'

4th grade score and the average 4th grade score was not enough to control for selection and reverse causation. On the other hand, because grading standards are estimated with error, and the measurement error is amplified when incorporating fixed effects, the decrease in the effect of grading standards could be attributed to attenuation bias.

The estimated effect using instrumental variables is lower at the school level but higher at the school subject level, especially when I incorporate fixed effects. This is consistent with the hypothesis that the OLS regressions suffered from attenuation bias, since the instrument also deals with measurement error.

To see the magnitude of the estimated effects, I calculate the effect of increasing standards in one standard deviation. The effect goes from 15.0 to 19.3 points, which corresponds to 0.30-0.38 standard deviations. To put this effect in perspective, the gender gap in test scores is 0.05 standard deviations, the urban-rural gap is 0.31 standard deviations, and the municipal versus voucher versus private school gap is 0.41 and 0.83 standard deviations, respectively.

Following Betts and Grogger (2003), another way of interpreting the results is to compare the predicted gains in students' test scores with the average gains in students' test scores. In this case, there is an average loss of -0.90 points between grades 4 and 8. The predicted effect of decreasing the grading standards by one standard deviation is more than 10 times larger than the mean loss between grades 4 and 8. Thus, the effect is substantial in terms of the mean difference.

A third way to interpret the results is to compare the predicted gain in students' tests scores with the average difference between the student's best test score and worst test score. The average difference is 41.58 points, so the effect of increasing grading standards by one standard deviation corresponds to 36-46% of the average difference.

The effect found in this study is larger than the one found by Betts and Grogger (2003). They found that the effect of increasing grading standards by one standard deviation increased test scores by less than one tenth of a standard deviation. The difference between the results in this study and the results in Betts and Grogger (2003) could have various explanations. One possible explanation is that they use data for 12th grade while I use data for 8th grade. If the effect of grading standards on test scores varies by age, then we would expect a difference in the results. To test this hypothesis, I would need to replicate this study for high school students. This is left for future research.

Because the theory predicts that increasing grading standards may result in some students giving up and thus reducing their levels of effort, I also present the results of quantile regression estimates. Columns (1)-(3) of Table 15 show results for the 25th, 50th and 75th percentiles of the 8th grade score. The effect of grading standards is positive and statistically significant for each of these percentiles. The results show that the effect of grading standards declines as the percentile goes up, and the difference between percentiles is statistically significant.

To estimate quantile regressions using instrumental variables, I follow the method proposed by Chernozhukov and Hansen (2008). Columns (4)-(6) in Table 15 present results for the instrumental variable quantile regressions for the 25th, 50th and 75th percentiles. Compared with the quantile regressions, the effects are smaller for all percentiles. Also, the difference between percentiles is statistically significant only between the 25th and 50th percentiles, but not between the 50th and 75th percentiles.

Overall, the evidence presented in Table 15 shows that the effect of grading standards is

Table 15: Estimated effects of grading standards on test scores: Quantile regressions

	Quantile Regressions			IV Quantile Regressions		
Variable	q25	q50	q75	q25	q50	q75
Standard(1)	0.733***	0.684***	0.654***	0.671***	0.642***	0.641***
	(0.0047)	(0.0049)	(0.0042)	(0.0056)	(0.0052)	(0.0056)
Observations	350,279	350,279	350,279	350,279	350,279	350,279

Notes: Standard errors are in parentheses. All models include subject fixed effects, controls for past test score and the past mean score, controls for mother and father's years of schooling, household income, an indicator for male student, indicators for urban schools, private schools and municipal schools. Standard(1) refers to grading standard measured at the school level.

stronger for low achieving students. This result is different from what Betts and Grogger (2003) and Figlio and Lucas (2004) found, in that both studies conclude that the effect is greatest for the top of the test distribution. Whether this is due to differences in how students respond to incentives in each country or differences in the grading system used in each country is left for future work.

Although the distributional effect is different from what was found in previous studies, this is still consistent with the theory: for an S-shaped utility function the theory predicts a negative effect for the lower tail of the ability distribution, but conditional on the effect being positive, the theory predicts that the effect will decline with ability.

2.5. Specification Checks

2.5.1. Differences in Curriculum Coverage

Instead of capturing difference in grading difficulty, the estimated standards may be capturing differences in curricula. For example, if students in school 1 are learning more advanced math than students in school 2, then, conditional on his math grade, a student in school 1 should get a higher standardized score, which would be an alternative explanation for my findings.

In Chile, the Ministry of Education prepares the National Minimum Curriculum (NMC). After covering it, schools have freedom of curriculum. However, in practice the minimum curriculum is so ambitious that very few schools are able to completely cover it. For example, in 2009, less than 1% of schools completely covered the NMC (Table 10).

In 2009, the SIMCE teacher's questionnaire asked teachers' own assessment of curriculum coverage by subject. Teachers were asked about 18 topics in language, 22 topics in mathematics and 30 topics in science. Teachers responded whether they had covered the topic completely (4), mostly covered it (3), covered something (2), or not covered it at all (1). I use these responses to construct the mean coverage by subject.

I estimate equation (24) and equation (25), controlling for the mean curriculum coverage by subject and an indicator variable for whether the NMC was completely covered for that subject. Then I use these new estimated standards and reestimate the models presented in Table 14. Results are presented in Table 16. The effect of grading standards on students' test scores is virtually unchanged after controlling for curriculum coverage.

Table 16: Estimated effects of grading standards on test scores controlling by curriculum

	OLS Regressions				IV Regressions			
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard(1b)	0.702***				0.657***			
	(0.0102)				(0.0193)			
Standard(2b)		0.612***	0.500***	0.492***		0.627***	0.566***	0.569***
		(0.0077)	(0.0081)	(0.0100)		(0.0166)	(0.0208)	(0.0203)
Observations	350,279	350,279	350,279	350,279	345,485	345,485	345,485	345,485
R-squared	0.569	0.575	0.338	0.84	0.569	0.576	0.037	0.075
School FE			yes				yes	
Student FE				yes				yes

Notes: Standard errors clustered at the school level are in parentheses. All models include subject fixed effects. (1), (2), (5) and (6) include controls for past test score and the past mean score. (1)-(3) and (5)-(7) include controls for mother and father's years of schooling, household income, and an indicator for male student. (1), (2), (5) and (7) include indicators for urban schools, private schools and municipal schools. Standard(1b) refers to the grading standard, controlling for curriculum coverage, measured at the school level, while Standard(2b) refers to the grading standard, controlling for curriculum coverage, measured at the school-subject level.

2.5.2. Nonlinearities in the Effect

To allow for possible nonlinearities in the effect of grading standards on students' tests scores, I estimate a simple variant of equations where the estimated standard is interacted with an indicator of whether the grading standard is among the lowest third, middle third or highest third of the estimated standard distribution.

The results presented in Table 17 show that we cannot reject that the effect of the standard is nonlinear. The average effect is significantly higher for the 33% schools (schools-subject) with the highest standards than it is for the 33% with the lowest standards. This is surprising as one would expect the effect to decrease as the standard increases, because of more and more students giving up[8].

2.5.3. Different Effect by Subject

Equations (27) and (28) assume that the effect of the standard on test scores is the same for every subject. We can relax the assumption and estimate the following equations:

$$S_{ijs} = \alpha + \gamma_j Standard_{js} + \delta_j + \xi_s + \upsilon_{ijs} \quad (30)$$

$$S_{ijs} = \alpha_i + \gamma_j Standard_{js} + \delta_j + \nu_{ijs} \quad (31)$$

The results are presented in Table 29. The effect is positive and statistically significant for all subjects, but the F-test rejects the hypothesis that the effect is the same for all subjects. The effect is strongest for math, then science and finally language. This is consistent with a nonlinear effect of standards on scores, in that the mean standard is lowest for language and highest for math.

[8]However, this result is still consistent with the theory because, conditional on $e > 0$, the sign of $\frac{d^2e}{d\tau^2}$ is ambiguous

Table 17: Estimated effects of grading standards on test scores: Exploring nonlinearities

		OLS Regressions				IV Regressions			
Variable		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard(1)	Easy	0.666***				0.567***			
		(0.0130)				(0.0209)			
	Med	0.723***				0.452***			
		(0.0300)				(0.0767)			
	Hard	0.741***				0.690***			
		(0.0136)				(0.0212)			
Standard(2)	Easy		0.542***	0.409***	0.406***		0.491***	0.406***	0.410***
			(0.0103)	(0.0101)	(0.0124)		(0.0174)	(0.0289)	(0.0253)
	Med		0.619***	0.501***	0.502***		0.573***	0.717***	0.740***
			(0.0227)	(0.0208)	(0.0255)		(0.0793)	(0.0832)	(0.0804)
	Hard		0.675***	0.573***	0.574***		0.680***	0.649***	0.656***
			(0.0104)	(0.0111)	(0.0138)		(0.0195)	(0.0245)	(0.0241)
Observations		350,279	350,279	350,279	350,279	350,279	350,279	350,279	350,279
R-squared		0.569	0.576	0.338	0.841	0.568	0.576	0.037	0.077
F test		9.46	46.91	69.80	44.27	40.41	93.30	68.09	79.09
School FE				yes				yes	
Student FE					yes				yes

Notes: Standard errors clustered at the school level are in parentheses. All models include subject fixed effects. (1), (2), (5) and (6) include controls for past test score and the past mean score. (1)-(3) and (5)-(7) include controls for mother and father's years of schooling, income, and an indicator for male student. (1), (2), (5) and (7) include indicators for urban schools, private schools and municipal schools. Standard(1) refers to the grading standard measured at the school level, while Standard(2) refers to the grading standard measured at the school-subject level.

Table 18: Estimated effects of grading standards on test scores: Effect by Subject

		OLS Regressions			IV Regressions		
Variable		(1)	(2)	(3)	(4)	(5)	(6)
Standard(2)	Math	0.645***	0.525***	0.523***	0.859***	0.568***	0.571***
		(0.0081)	(0.0083)	(0.0103)	(0.0112)	(0.0216)	(0.0204)
	Language	0.601***	0.426***	0.421***	0.815***	0.448***	0.450***
		(0.0095)	(0.0099)	(0.0121)	(0.0141)	(0.0279)	(0.0265)
	Science	0.600***	0.477***	0.477***	0.818***	0.520***	0.527***
		(0.0084)	(0.0085)	(0.0104)	(0.0105)	(0.0226)	(0.0215)
Observations		350,279	350,279	350,279	350,279	350,279	350,279
R-squared		0.576	0.338	0.841	0.302	0.038	0.079
F test		36.04	130.56	92.16	89.85	159.02	178.91
School FE			yes			yes	
Student FE				yes			yes

Notes: Standard errors clustered at the school level are in parentheses. All models include subject fixed effects. (1) and (4) include controls for past test score and the past mean score. (1), (2), (4) and (5) include controls for mother and father's years of schooling, household income, and an indicator for male student. (1) and (4) include indicators for urban schools, private schools and municipal schools. Standard(2) refers to the grading standard measured at the school-subject level.

2.6. Conclusions

This study identifies one effective teaching practice: setting high grading standards. My results provide new evidence that grading standards do have a significant effect on academic performance. Students' tests scores rise in schools and subjects with higher standards. Moreover, the evidence suggests that increasing grading standards has desirable distributional consequences, because the effect is stronger for the bottom of the achievement distribution.

It is tempting to conclude from the results of this study that the policy implication is to increase grading standards. However, we don't know what the effects of high grading standards would be in high school. One important college admission requirement in Chile is high school GPA, so a teacher who uses high grading standards would lower the probability of her students getting into college relative to a teacher with low grading standards. A careful analysis of grading standards at the secondary level is left for future work.

3. Does teacher gender matter?

3.1. Introduction

Numerous studies have documented a gap in the average educational achievement between boys and girls. This gap is especially important in math and language, with boys outperforming girls in math and girls outperforming boys in language. Table 19 presents evidence of the gender gap in Chile, but this gap is also present in developed countries such as the United States, Australia and England (Mead, 2006). As the table shows, there is a gap of approximately 9 points in math and 10 points in language.

It is important to understand the factors determining this gap, as it may drive gender differences in the labor market. For example, women in Chile tend to study fields leading to careers in education and health, whereas men tend to study fields leading to careers in science and math, which on average are associated with higher wages. This may have implications for women's returns to schooling and may relate to occupational segregation and earnings inequality by gender (Loury, 1997). It is important to analyze the factors that determine the gap because it is related to the economic equality for women.

One explanation that has been discussed in the literature emphasizes the gender of language and math teachers. First, the gender of the teacher can have an effect on students' behavior through role model effects or through stereotype threats (Dee, 2007). If we think of teachers as role models, and if students identify themselves more with same-sex role models, then it is possible that performance will be enhanced when students are assigned to a same gender teacher. The same result is also consistent with the theory of stereotype threats, which states that, in the case of negative stereotypes against a group, group members may internalize the stereotypes as explanations of their own behavior (Holmlund and Sund, 2008). In both cases, it is the student who is reacting to the gender of the teacher.

Second, the teacher gender may matter because of teachers' behavior. For example, teachers might have a preference towards students of their own sex, and hence female (male) teachers will structure their classrooms in ways that enhance girls' (boys') learning. If not preferences, gender stereotypes may influence teachers' behavior. In both cases, it is the teacher reacting to the gender of the student.

The previous literature has found mixed results for the effect of female teachers. Dee (2007) finds that assigning an opposite gender teacher lowers student achievement. His study focuses on differences in teacher perceptions of student behavior. Ammermuller and Dolton (2006), using the same methodology as Dee (2007), show that in England and the United States, students perform better when their teachers are of the same gender.

Nixon and Robinson (1999) estimate the effect of the percentage of high school female faculty on female years of schooling, high school graduation, enrollment in college and graduation from college. They find a positive effect of female faculty on female students, while no effect on male students. Bettinger and Long (2005) show that the presence of faculty members of the same gender has a positive and significant impact on course selection and on choice of major. On the other hand, in a study of Latin America, Ramirez and Ursini (2005) find the opposite result. Using teacher interviews and class observations, they find that female teachers consider mathematics to be a male domain, so the type of interaction they have with girls and boys favors the latter.

Table 19: Gender Gap in SIMCE 2009: Scores by Subject

	Boys	Girls	Difference
Total	256.61	254.26	2.35
	(52.58)	(50.02)	(0.11)
	455958	455315	
Math	264.06	255.19	8.87
	(52.15)	(51.34)	(0.22)
	114088	113905	
Lang	246.52	257.11	-10.59
	(51.60)	(50.12)	(0.21)
	113933	113715	
Natural Sc	261.45	256.42	5.03
	(52.70)	(49.37)	(0.21)
	114061	113854	
Social Sc	254.39	248.34	6.05
	(52.10)	(48.74)	(0.21)
	113876	113841	

Notes: Standard errors are presented in parentheses.
Number of observations are presented in italics.

In this essay I investigate the effect of the teacher gender on the educational achievement of boys and girls. For this purpose, I measure the educational performance as the result of a set of standardized tests called SIMCE. The SIMCE consists of four tests: Language and Communication (Lang), Mathematical Education (Math), Study and Comprehension of Nature (Natural Sc) and Study and Comprehension of Society (Social Sc), where knowledge learned recently as well as past learning is measured. 8th grade students in Chile take this test at the end of the school year.

Sections 3.2 and 3.3 introduce the data and econometric framework used in this essay. In Section 3.4 and 3.5 the main results are analyzed and the internal validity of the estimate is discussed. Section 3.6 presents evidence regarding the possible mechanisms, and Section 3.7 concludes.

3.2. Data

The data sets I use are the SIMCE 2007 and SIMCE 2009, as well as data from the Ministry of Education (MINEDUC). The SIMCE provides census information on standardized educational performance tests, characteristics of the household and characteristics of the school and teachers. In 2007, SIMCE evaluated 263,913 students registered in 8th grade, and 5,806 schools (MINEDUC 2008). This covers 96% of students and 99% of schools. In 2009, SIMCE evaluated 239,745 students registered in 8th grade, and 5,814 schools (MINEDUC 2010), covering 92.5% of students and 98.2% of schools.

The SIMCE test, created in 1998, is the main instrument with which the quality of education in Chile is measured. In 1999, the test was modified and standardized in order to follow up on school performance. The new test has an open-ended scale that measures student abilities (cognitive skills). It uses the Item Response Theory (IRT) which is applied in most international tests on academic achievement (such as TIMSS or PISA) and links students' score to their abilities. In 1999, it was also determined that

Table 20: Data: Sample versus Population

	Population	Sample	Difference
SIMCE	254.78	259.70	4.91
	(50.94)	(50.65)	(0.07)
	1934717	*675192*	
Math	257.49	261.92	4.43
	(51.80)	(51.54)	(0.14)
	484119	*197613*	
Language	252.36	257.04	4.68
	(51.11)	(50.67)	(0.13)
	483103	*227873*	
Nat. Sciences	258.38	263.25	4.87
	(50.35)	(49.78)	(0.15)
	483869	*134011*	
Soc. Sciences	250.90	257.02	6.12
	(50.07)	(49.64)	(0.16)
	483626	*115695*	

Notes: Standard errors are presented in parentheses.
Number of observations are presented in italics.

the mean for that year would be 250, with a standard deviation of 50. The parameters of all subsequent evaluations have been calculated on the basis of the mean and variance of 1999.

Eighth graders take four tests: Mathematics, Language and Communication, Social Sciences and Natural Sciences. The scores for these four tests and socioeconomic characteristics of the students' households are available for 255,012 students. In addition, SIMCE has information on the teachers of these four subjects. Four complete teacher surveys are available for 274,323 students. The teacher survey has information on teacher background and information on the material covered in class.

I also use data from MINEDUC that has information on grades per subject for the years 2008 and 2006. Data on grades is available for 274,686 students. For the purpose of this study, I standardize the scores and grades by subject.

For this study, I use a sample of students with valid SIMCE scores, household characteristics and past grades, and for whom the teacher survey is available. This information is available for 221,273 students. Table 20 shows how the sample used in this study compares to the whole population. My sample has higher SIMCE scores, which could compromise the external validity of my results. I will return to this issue after presenting the results.

3.3. Econometric Framework

To estimate the effect of a female teacher on students' test scores, I start by estimating the following equation:

$$y_{ijcst} = \alpha_s + \gamma_t + \delta_j + X'_{ijcst}\lambda + FT_{jcst}\Pi + Z'_{jcst}\psi + \varepsilon_{ijcst} \qquad (32)$$

where y_{ijsct} is the standardized test score for student i in subject j, in classroom c, in school s, at period t. α_s is a school fixed effect, X_{ijsct} is a vector of student characteristics,

Table 21: Gender Gaps by Subject

	Girls	Boys	Difference
SIMCE	-0.02	0.02	0.04
(Standardized)	(0.98)	(1.02)	(0.00)
	342346	*332846*	
Math	-0.10	0.11	0.21
	(0.99)	(1.00)	(0.00)
	100507	*97106*	
Language	0.11	-0.11	-0.22
	(0.97)	(1.02)	(0.00)
	115039	*112834*	
Nat. Sciences	-0.08	0.08	0.15
	(0.97)	(1.02)	(0.01)
	67708	*66303*	
Soc. Sciences	-0.07	0.07	0.14
	(0.97)	(1.02)	(0.01)
	59092	*56603*	

Notes: Standard errors are presented in parentheses. Number of observations are presented in italics. SIMCE is standardized to have a mean of 0 and standard deviation of 1.

Table 22: Teacher Characteristics by Gender

	Female	Male	Difference
Experience	19.8840	20.9961	-1.1120
	(12.1712)	(12.5917)	(0.0330)
Certification	0.9877	0.9723	0.0155
	(0.1100)	(0.1643)	(0.0003)
Specialization	0.6810	0.6469	0.0341
	(0.4661)	(0.4779)	(0.0013)

Notes: Standard errors are presented in parentheses.

including mother and father's education and household income, Z_{jcst} is a vector of teacher characteristics, including experience, certification and specialization, γ_t is a year fixed effect, δ_j is a subject fixed effect and FT_{jcst} is a dummy variable for a female teacher.

The main problem when studying the effect of the teacher on student performance using equation (32) is that the results may be biased by the non-random assignment of students to teachers. For example, higher achieving students could be assigned to female teachers. In addition, there could be gender-specific patterns in teachers' classroom assignment.

To overcome these difficulties, Dee (2007) proposes to estimate the following equation:

$$y_{ijcst} = \mu_i + \gamma_t + \delta_j + X'_{ijcst}\lambda + FT_{jcst}\Pi + Z'_{jcst}\psi + \nu_{ijcst} \tag{33}$$

where μ_i is a student fixed effect. The estimation of equation (33) is possible only if we have multiple observations per student. Dee (2007) has two observations for the same student in different subjects, so he estimates equation (33) using first differences. In my case, I observe each student in four different subjects.

As Dee (2007) argues, the main threat to identification when estimating equation (33) is that assignment to a female teacher may depend on students' subject-specific propensity for achievement. Also, unobserved teacher and classroom traits could be correlated with gender.

The advantage of my data set is that I can control for subject specific propensity for achievement using students' previous year grades by subject. Then I can estimate the following equation:

$$y_{ijcst} = \mu_i + \gamma_t + \delta_j + X'_{ijcst}\lambda + FT_{jcst}\Pi + Z'_{jcst}\psi + G_{ijc-1s-1t-1}\phi + \nu_{ijcst} \tag{34}$$

where $G_{ijc-1s-1t-1}$ are the student's grades in the previous year.

However, the estimated effect may be capturing unobserved determinants of teacher quality and classroom traits. These concerns are discussed in Section 3.5

3.4. Results

Table 23 shows the results for the effect of being assigned to a female teacher. Columns (1) to (3) estimate the effect using school fixed effects, while columns (4) to (6) use individual fixed effects. All specifications control for parental income and education, and for subject fixed effects. Models (2), (3), (5) and (6) include teacher controls, such as experience, certification and specialization. Models (3) and (6) control for the student's past grades.

When controlling for school fixed effects, the effect of being assigned to a female teacher is positive and statistically significant for both girls and boys. However, when controlling for individual fixed effects, the effect is no longer statistically significant for boys. The comparison between models with student fixed effects and school fixed effects suggests that the propensity for girls to be assigned to a female teacher is unrelated to the unobserved determinants of achievement. However, the results for boys suggest that boys with an unobserved propensity for low achievement are less likely to be assigned to female teachers.

The estimated effect of being assigned to a female teacher is reduced in models that introduce teacher and past grades controls, but the reduction is small and not statistically

Table 23: Estimated effects of female teachers on test scores

	Boys					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Female teacher	0.0157***	0.0157***	0.0136**	0.00137	-0.00045	0.00118
	(0.0050)	(0.0051)	(0.0056)	(0.0059)	(0.0060)	(0.0064)
Past grade			0.437***			0.266***
			(0.0026)			(0.0042)
Observations	332,846	332,846	332,846	332,846	332,846	332,846
R-squared	0.309	0.309	0.469	0.761	0.761	0.780
Prob > F		0.148	0.032		0.000	0.000
	Girls					
	(1)	(2)	(3)	(4)	(5)	(6)
Female teacher	0.0440***	0.0438***	0.0378***	0.0443***	0.0425***	0.0395***
	(0.0050)	(0.0050)	(0.0061)	(0.0063)	(0.0063)	(0.0069)
Past grade			0.438***			0.251***
			(0.0026)			(0.0044)
Observations	342,346	342,346	342,346	342,346	342,346	342,346
R-squared	0.321	0.321	0.491	0.760	0.760	0.777
Prob > F		0.098	0.006		0.000	0.000
School Fixed Effect	yes	yes	yes			
Student Fixed Effect				yes	yes	yes
Teacher Controls		yes	yes		yes	yes

Notes: Standard errors, clustered at the school level, are presented in parentheses. All models include controls for parental income and schooling, and for subject and time fixed effects. Prob>F refers to an F-test of the joint significance of the teacher controls.

significant.

3.5. Robustness Checks

As discussed in Section 3.3, the estimated effect of being assigned to a female teacher could be capturing unobserved determinants of teacher quality and classroom traits. Table 22 shows that male and female teachers differ in observable characteristics. Female teachers are on average less experienced than their male counterparts, but have more certification and specializations. These controls are jointly significant, and, if excluded, the effect of female teachers is overestimated. It is then possible that unobserved teacher characteristics are biasing the results in the same direction.

To control for unobserved teacher quality, I would need a panel where I could observe the same teacher for multiple years. Unfortunately such data is not available. However, the effect of a female teacher is positive only for girls. This suggests either that unobserved teacher quality is not biasing the results, or, alternatively, it could be overestimating the positive results for girls and hiding a negative effect for boys.

To explore this possibility, I estimate the following equation:

$$y_{ijcst} = \kappa_j + X'_{ijcst}\lambda + OppGender_{jcst}\Pi + \varepsilon_{ijcst} \tag{35}$$

where the variable *OppGender* is equal to 1 if the student and teacher are from the opposite gender, and κ_j is a teacher fixed effect.

Table 24: Estimated effects of female teachers on test scores: Teacher Fixed Effects

VARIABLES	(1)	(2)
OppGender	-0.0520***	-0.0347***
	(0.00315)	(0.00258)
Past Grade		0.481***
		(0.00223)
Observations	675,192	675,192
R-squared	0.359	0.535

Notes: Standard errors, clustered at the school level, are presented in parentheses. All models include controls for parental income and schooling, and for subject fixed effects.

Results of model (35) are presented on Table 24. The results show that a female teacher improves the score of girls by 3.5% of a standard deviation. This is slightly lower than the effect presented on Table 23.

The disadvantage of estimating equation (35) is that it controls for teacher unobserved characteristics at the expense of assuming that the effect of the gender of the teacher is symmetric: girls assigned to a female teacher have on average the same increase in test scores as boys assigned to a male teacher. If the symmetry assumption does not hold, the results could be biased. A data set with multiple years of observations per teacher is needed to address these concerns.

The concern about classroom unobserved traits biasing the results is not an issue in the models using student fixed effects, because in Chile students are assigned to the same classroom for every subject. To test if unobserved classroom traits bias the results in the models with school fixed effects, I reestimate the model using classroom fixed effects. Results are presented in Table 25. Compared to the results of the estimation with school fixed effects, the coefficients in the estimation with classroom fixed effects are lower. This indeed suggests that female teachers are being assigned to classrooms with better classroom traits.

Equation 33 assumed that only variables in t (with the exception of grades) have an effect on student's test scores. Alternative, we could allow the gender of teachers in previous years to have an effect on test scores:

$$y_{ijcst} = \mu_i + \gamma_t + \delta_j + X'_{ijcst}\lambda + \sum_{\tau=1}^{t} FT_{jcs\tau}\Pi_\tau + Z'_{jcst}\psi + G_{ijc-1s-1t-1}\phi + \nu_{ijcst} \tag{36}$$

The data set I am using does not have observations on the history of teachers, but I can construct the predicted probability of being assigned to a female teacher using the estimated coefficients from the following equation:

$$Prob(FT_{ijcst} = 1) = \alpha_s + \delta_j + \beta_1 Male_i + \beta_2 G_{ijc-1s-1t-1} + \beta_3 \%FT_{jst} + \varepsilon_{ijcst} \tag{37}$$

Table 25: Estimated effects of female teachers on test scores: Classroom Fixed Effects

	Boys			Girls		
VARIABLES	(1)	(2)	(3)	(1)	(2)	(3)
Female teacher	0.0140***	0.0138***	0.0130**	0.0417***	0.0411***	0.0356***
	(0.00454)	(0.00453)	(0.00586)	(0.00482)	(0.00482)	(0.00657)
Past grade			0.437***			0.440***
			(0.00273)			(0.00266)
Observations	332,846	332,846	332,846	342,346	342,346	342,346
R-squared	0.363	0.363	0.510	0.374	0.374	0.530
F-test		1.536	2.235		1.401	2.506
Prob >F		0.104	0.008		0.157	0.003
Classroom Fixed Effect	yes	yes	yes	yes	yes	yes
Teacher Controls		yes	yes		yes	yes

Notes: Standard errors, clustered at the school level, are presented in parentheses. All models include controls for parental income and schooling, and subject fixed effects. Prob>F refers to an F-test of the joint significance of the teacher controls.

Table 26: Probability of Being Assigned to a Female Teacher

VARIABLES	(1)
male	-0.00512***
	(0.0008)
Past Grade	0.00041
	(0.0004)
% Female	0.645***
	(0.0015)
Observations	636,731
R-squared	0.529

Notes: Standard errors in parentheses. Model includes subject and school fixed effects.

where $\%FT_{jst}$ is the percentage of female teachers by subject in the school. Results are presented in Table 26, and the estimated coefficients are used to construct $\hat{FT}_{ijc-1s-1t-1} = \hat{Prob}(FT_{ijc-1s-1t-1} = 1)$. Then I can estimate the following model:

$$y_{ijcst} = \mu_i + \gamma_t + \delta_j + X'_{ijcst}\lambda + FT_{jcs\tau}\Pi_t + \hat{FT}_{jc-1s-1t-1}\Pi_{t-1} + Z'_{jcst}\psi + G_{ijc-1s-1t-1}\phi + \nu_{ijcst} \tag{38}$$

Table 27 estimates the effect on test scores of being assigned to a female teacher, controlling for female teachers by subject in the past period. When controlling for past exposure, the effect of being assigned to a female teacher is reduced for girls, but the reduction is small. Consistent with my previous results, past exposure has a positive effect for girls and a negative effect for boys, although the effect is insignificant for both girls and boys.

Table 27: Estimated effects of female teachers on test scores

				Boys		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Female teacher	0.0123**	0.0121**	0.0117*	0.00754	0.00615	0.00643
	(0.00616)	(0.00615)	(0.00652)	(0.00651)	(0.00656)	(0.00713)
Female teacher	0.0255**	0.0249**	0.0158	-0.0117	-0.0125	-0.00865
in t-1	(0.01080)	(0.01080)	(0.01180)	(0.01390)	(0.01390)	(0.01420)
Past grade			0.439***			0.263***
			(0.00277)			(0.00446)
Observations	309,919	309,919	309,919	309,919	309,919	309,919
R-squared	0.313	0.314	0.474	0.775	0.775	0.792
Prob > F		0.063	0.021		0.000	0.000
				Girls		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Female teacher	0.0400***	0.0400***	0.0376***	0.0387***	0.0373***	0.0366***
	(0.00583)	(0.00585)	(0.00674)	(0.00690)	(0.00690)	(0.00758)
Female teacher	0.0202*	0.0199*	0.0031	0.0220	0.0215	0.0140
in t-1	(0.01070)	(0.01070)	(0.01200)	(0.01640)	(0.01640)	(0.01730)
Past grade			0.440***			0.249***
			(0.00277)			(0.00465)
Observations	320,902	320,902	320,902	320,902	320,902	320,902
R-squared	0.325	0.325	0.496	0.773	0.773	0.789
Prob > F		0.068	0.003		0.000	0.000
School Fixed Effect	yes	yes	yes			
Student Fixed Effect				yes	yes	yes
Teacher Controls		yes	yes		yes	yes

Notes: Standard errors, clustered at the school level, are presented in parentheses. All models include controls for parental income and schooling, and subject and time fixed effects. Prob>F refers to an F-test of the joint significance of the teacher controls.

Table 28: Percentage of Female Teachers by Subject

	Math	Lang	Natural Sc	Soc Sc
% female teacher	0.5853	0.8293	0.7553	0.6167
	(0.0056)	(0.0039)	(0.0058)	(0.0071)

Notes: Standard errors are presented in parentheses

3.6. Possible Mechanisms

As discussed in Section 3.1, the positive effect of the female teacher on girls may have different explanations. If it is due to teachers providing role model or the theory of stereotype threats, we would expect a larger effect for girls in subjects that are considered male dominated: we should observe a larger effect of female teachers in math and a smaller effect in language. Also, we would expect to see a larger effect on girls who don't have other strong female role models. If the positive effect of the female teacher on girls is due to teacher behavior, rather than role modeling, then we won't necessarily see a different effect by subject or for girls without other strong female role models.

Equations (32) and (33) assume that the effect of being assigned to a female teacher is the same, independent of the subject being taught. Alternatively, and consistent with the role model hypothesis, a female teacher has a larger effect for girls in subjects that are considered male dominated. Table 28 shows the percentage of female teachers by subject. Although there are more female teachers in every subject, the percentage is lower in math and higher in language. This is consistent with the belief that math is a male dominated subject while language is a female dominated subject.

To test this hypothesis, I estimate the following equations:

$$y_{ijcst} = \alpha_s + \gamma_t + \delta_j + X'_{ijcst}\lambda + FT_{jcst}\Pi_j + Z'_{jcst}\psi + \varepsilon_{ijcst} \tag{39}$$

and

$$y_{ijcst} = \mu_i + \gamma_t + \delta_j + X'_{ijcst}\lambda + FT_{jcst}\Pi_j + Z'_{jcst}\psi + G_{ijc-1s-1t-1}\phi + \nu_{ijcst} \tag{40}$$

Results are presented in Table 29. For boys, the effect of being assigned to a female teacher is not significantly different from zero for any subject when using student fixed effects. For girls, the effect is positive for all subjects but only statistically significant for mathematics and social sciences. These are the subjects with a lower percentage of female teachers as shown in Table 28. However, the F test cannot reject that there is no difference in the effect by subject.

If the role model hypothesis is true, the effect should be greater for girls who are lacking other positive female role models. To test this, I divide the sample by mother's education and create an indicator of whether the mother has less than 8 years of education (no high school), between 8 and 12 (some high school to high school graduate), or 12 or more years (at least some years of college or professional school). Results are presented in Table 30. The F test rejects the hypothesis that the effect is the same for all levels of

Table 29: Estimated effects of female teachers on test scores by subject

				Boys		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Fem teacher	-0.00289	-0.00301	-0.00208	-0.00843	-0.0132	-0.00953
math	(0.00866)	(0.00871)	(0.00964)	(0.01090)	(0.01090)	(0.01160)
Fem teacher	0.00863	0.00836	0.0106	-0.0168	-0.0193*	-0.0138
language	(0.01030)	(0.01030)	(0.01200)	(0.01160)	(0.01150)	(0.01290)
Fem teacher	0.0261**	0.0262**	0.0183	0.0167	0.016	0.0152
natural sciences	(0.01090)	(0.01090)	(0.01330)	(0.01270)	(0.01280)	(0.01440)
Fem teacher	0.0450***	0.0453***	0.0386***	0.0210*	0.0234*	0.0202
social sciences	(0.01040)	(0.01040)	(0.01200)	(0.01220)	(0.01230)	(0.01360)
Past Grade			0.437***			0.266***
			(0.00264)			(0.00419)
Observations	332,846	332,846	332,846	332,846	332,846	332,846
R-squared	0.309	0.309	0.469	0.761	0.761	0.78
Prob > F	0.003	0.003	0.086	0.055	0.021	0.176
				Girls		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Fem teacher	0.0436***	0.0435***	0.0400***	0.0551***	0.0499***	0.0489***
math	(0.00823)	(0.00826)	(0.00907)	(0.01040)	(0.01050)	(0.01080)
Fem teacher	0.0413***	0.0409***	0.0383***	0.0223*	0.0226*	0.0217
language	(0.01050)	(0.01040)	(0.01220)	(0.01190)	(0.01210)	(0.01330)
Fem teacher	0.0359***	0.0358***	0.0137	0.0367***	0.0356***	0.0247*
natural sciences	(0.01070)	(0.01070)	(0.01360)	(0.01260)	(0.01270)	(0.01490)
Fem teacher	0.0541***	0.0542***	0.0545***	0.0525***	0.0538***	0.0527***
social sciences	(0.01020)	(0.01020)	(0.01260)	(0.01190)	(0.01200)	(0.01330)
Past Grade			0.438***			0.251***
			(0.00261)			(0.00435)
Observations	342,346	342,346	342,346	342,346	342,346	342,346
R-squared	0.321	0.321	0.491	0.76	0.76	0.777
Prob > F	0.654	0.644	0.157	0.131	0.203	0.175
School Fixed Effect	yes	yes	yes			
Student Fixed Effect				yes	yes	yes
Teacher Controls		yes	yes		yes	yes

Notes: Standard errors, clustered at the school level, are presented in parentheses. All models include controls for parental income and schooling, and subject and time fixed effects. Prob>F refers to an F-test of equal effect by subject.

Table 30: Estimated effects of female teachers on test scores by mother's level of schooling

VARIABLES	Boys (1)	Girls (1)
Fem teacher*(low mother educ)	0.0245***	0.0533***
	(0.00866)	(0.00865)
Fem teacher*(medium mother educ)	0.00355	0.0419***
	(0.00778)	(0.00799)
Fem teacher*(high mother educ)	-0.0134	0.0147
	(0.00959)	(0.01080)
Past Grade	0.250***	0.237***
	(0.00398)	(0.00405)
Observations	332,846	342,346
R-squared	0.789	0.788
Prob > F	0.006	0.011
Student Fixed Effect	yes	yes
Teacher Controls	yes	yes

Notes: Standard errors, clustered at the school level, are presented in parentheses. Model include controls for subject fixed effects. Prob >F refers to an F-test of equal effect by mother education

mother's education. Students whose mother has lower levels of education benefit more from having a female teacher than students whose mother has a high level of education. This holds for boys and girls, although the effect is always greater for girls than for boys.

3.7. Conclusions

In this essay, data from the Chilean SIMCE test has been used to empirically estimate a reduced form equation, shedding some light on one of the components of the gender gap for school students. The results show that there is an effect of the gender of the teacher on the scores of female students. Having a female math teacher increases the average scores of female students in the SIMCE test by approximately 0.04 standard deviations, which is almost one fourth of the gender gap in math and language. However, there is no evidence of an effect of teacher gender on boys' scores.

My results differ from those previously found by Dee (2007) because I do not find an effect for male students. They also differ from the results for Mexico found by Ramirez and Ursini (2005). My results are closer to those of Nixon and Robinson (1999), who find that a higher proportion of female faculty has a positive effect on females attending high schools, while there is no effect on male students.

The evidence is consistent with the role model hypothesis, as the effect is only significant for subjects with lower proportions of female teachers, and only for students whose mother has low levels of education.

Table 20 shows that the sample used in this study has students with lower SIMCE scores than the whole population. Because parental education is a predictor of SIMCE scores, the results for the sample should be lower than the corresponding results for the whole population.

References

[1] Ammermuller, A., Dolton, P., 2006, "Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA," *ZEW Discussion Paper*, pp. 06-600.

[2] Angrist, J., Lavy, V., 2009, "The Effects of High-Stakes High School Achievement Awards: Evidence from a Group-Randomized Trial," *American Economic Review*, vol. 99, no. 4, pp. 1384-1414.

[3] Angrist, J., Lang, D., Oreopoulos, P., 2009, "Incentives and Services for College Achievement: Evidence from a Randomized Trial," *American Economic Journal: Applied Economics*, American Economic Association, vol. 1, no. 1, pp. 136-63.

[4] Athey, S., 2004, "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information," *Econometrica*, vol. 69, no. 4, pp. 861-889.

[5] Babcock, P., 2009, "Real Costs of Nominal Grade Inflation? New Evidence from Student Course Evaluations," *Departmental Working Papers*, Department of Economics, UCSB, UC Santa Barbara

[6] Barut, Y., Kovenock, D., Noussair, C.N., 2002, "A Comparison of Multiple-Unit All-Pay and Winner-Pay Auctions under Incomplete Information," *International Economic Review*, vol. 43, no. 3, pp. 675-708

[7] Baye, M.R., Kovenock, D., de Vries, C.G., 1993, "Rigging the lobbying process: An application of the all-pay auction," *American Economic Review*, vol. 83, no. 1, pp. 289-294.

[8] Becker, W.E., Rosen, S., 1992, "The Learning effect of Assessment and Evaluation in High School," *Economics of Education Review*, vol. 11, no. 2, pp. 107-118.

[9] Bettinger, P., Long, B., 2005, "Do Faculty serve as a Role Model?," *American Economic Review*, vol. 95, no. 2.

[10] Betts, J., 1995, "Do grading standards affect the incentive to learn?" Manuscript, University of California, San Diego, Department of Economics.

[11] Betts, J., 1998, "The Impact of Educational Standards on the Level and Distribution of Earnings," *The American Economic Review*, vol. 88, no. 1, pp. 266-275.

[12] Betts, J., Grogger, J., 2003, "The impact of grading standards on student achievement, educational attainment, and entry-level earnings," *Economics of Education Review*, vol. 22, no. 4, pp. 343-352.

[13] Boyd, D., Grossman P., Lankford, H., Loeb, S., Wyckoff, J., 2006, "How changes in entry requirements alter the teacher workforce and affect student achievement," *Education Finance and Policy*, vol 1, no. 2, pp. 176-216.

[14] Chernozhukov and Hansen, 2008, "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics*, vol. 142, pp. 379-398.

[15] Clark, D.J., Riis, C., 1998, "Competition over more than one Prize," *American Economic Review*, vol. 88, pp. 276-289.

[16] Costrell, R., 1994, "A Simple Model of Educational Standards," *American Economic Review*, vol. 84, no. 4, pp. 956-71.

[17] Dee, T., 2007, "Teachers and the gender gaps in student achievement," *Journal of Human Resources*, vol. 42, pp. 528-554.

[18] DiNardo, J., Fortin, N. and Lemieux, T., 1996, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, vol. 64, no. 5, pp. 1001-44.

[19] Dubey, P., Geanakoplos, J., 2010, "Grading exams: 100, 99, 98, . . . or A, B, C?," *Games and Economic Behavior*, vol. 69, pp. 72-94.

[20] Figlio, D., Lucas, M., 2004, "Do high grading standards affect student performance?," *Journal of Public Economics*, vol. 88, no. 9-10, pp.1815-1834.

[21] Green, J. R., Stokey, N.L., 1983, "A comparison of tournaments and contracts," *Journal of Political Economy*, vol. 91, no. 3, pp. 349-364.

[22] Hanushek, E., Rivkin, S., 2004, "How to improve the supply of high quality teachers," *In Brookings Papers on Education Policy 2004, edutes by Diane Ravith. Washington DC: Brookings Institution Press*, pp. 7-25.

[23] Hanushek, E., Rivkin, S., 2006, "Teacher Quality," *In Handbook of Economics of Education, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland*, pp. 1051-1078.

[24] Hanushek, E., Rivkin, S., 2010, "Generalizations about using value-added measures of teacher quality," *American Economic Review*, vol. 100, no. 2.

[25] Harris, D., Sass, T., 2006, "Value-Added Models and the Measurement of Teacher Quality," *Florida State University*,

[26] Holmlund, H., Sund, K., 2008, "Is the gender gap in school performance affected by the sex of the teacher," Labour Economics, Elsevier, vol. 15, no. 1, pp. 37-53.

[27] Kane, T., Rockoff, J., Staiger, D., 2008, "What does certification tell us about teacher effectiveness? Evidence from New York City," *Economics of Education Review*, vol. 27, no. 6, pp. 615-631.

[28] Kane, T., Taylor, E., Tyler, J., Wooten, A., 2011, "Identifying Effective Classroom Practices Using Student Achievement Data," *Journal of Human Resources*, pp. 587-613.

[29] Kang, S., 1985, "A Formal Model of School Reward Systems, in John H. Bishop, ed., Incentives, learning, and employability," *Columbus, OH: National Center for Research in Vocational Education, Ohio State University*, pp. 27-38.

[30] Kremer, M., Miguel, E., Thornton, R., 2009, "Incentives to learn," *Review of Economics and Statistics*, no. 91, vol. 3.

[31] Lavy, V., 2011, "What Makes an Effective Teacher? Quasi-Experimental Evidence," *NBER Working Paper*, no. 16885.

[32] Lazear, E.P., Rosen, S., 1981, "Rank-order tournaments as optimum labor contracts," *Journal of Political Economy*, vol. 89, no. 5, pp. 841-864.

[33] Lewis, M., Lockhead, M., "Social Exclusion and the Gender Gap in Education," *The World Bank, Policy Research Working Paper Series*, no. 4562.

[34] Loury, L., 1997, "The Gender Earnings Gap among College-Educated Workers," *Industrial and Labor Relations Review*, vol. 50, no. 4, pp. 580-59.

[35] Mead, S., 2006, "The Truth about Boys and Girls," www.educationsector.org

[36] Majzub, R., Rais, M., 2010, "Boys' Underachievement: Male versus Female Teachers," *Procedia Social and Behavioral Sciences*, vol. 7(C), pp. 685-690.

[37] Milgrom, P., Shannon, C., 1994, "Monotone Comparative Statics," Econometrica, vol. 62, pp. 157-180.

[38] Nixon, L., Robinson, M., 1999, "The educational attainment of young women: Role model effects of female high school faculty," *Demography*, vol. 36, no. 2 .

[39] Ramirez, M., Ursini, S., 2005, "Influence of the female teachers' gender vision on the type of interaction they establish with boys and girls in the mathematics classroom,"

[40] Raymond, M., 2008, "Paying for A's: An Early Exploration of Student Reward and Incentive Programs in Charter Schools," *Stanford, CA: Stanford University, Center for Research on Educational Outcomes (CREDO).*

[41] Tyler, J., Taylor, E., Kane, T., Wooten, A., 2010, "Using Student Performance Data to identify Effective Teachers and Teaching Practices," *American Economic Review*, vol. 100, no. 2, pp. 256-260.

A. Appendix: Proofs of Results

Proof of Result 1.2
Define f as $f = v_K G(x_K) - S^* - \sigma_\theta x_K - v_K G(\frac{-S^*}{\sigma_\theta})$

$$\frac{\partial f}{\partial \sigma_\theta} = v_K g(x_K)\frac{\partial x_K}{\partial \sigma_\theta} - \frac{\partial S^*}{\partial \sigma_\theta} - \sigma_\theta \frac{\partial x_K}{\partial \sigma_\theta} - x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{-\frac{\partial S^*}{\partial \sigma_\theta}\sigma_\theta + S^*}{\sigma_\theta}$$

Using the definition of x_K

$$\begin{aligned}
\frac{\partial x_K}{\partial \sigma_\theta} &= \frac{1}{g'(x_K)}\frac{1}{v_K} \\
g(x_K) &= \frac{\sigma_\theta}{v_K}
\end{aligned}$$

where $\frac{\partial x_K}{\partial \sigma_\theta} < 0$ for the SOC to hold.

$$\begin{aligned}
\frac{\partial f}{\partial \sigma_\theta} &= v_K g(x_K)\frac{1}{g'(x_K)}\frac{1}{v_K} - \frac{\partial S^*}{\partial \sigma_\theta} - \sigma_\theta \frac{1}{g'(x_K)}\frac{1}{v_K} - x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\big(- \frac{\partial S^*}{\partial \sigma_\theta} + \frac{S^*}{\sigma_\theta}\big) \\
&= v_K \frac{\sigma_\theta}{v_K}\frac{1}{g'(x_K)}\frac{1}{v_K} - \frac{\partial S^*}{\partial \sigma_\theta} - \sigma_\theta \frac{1}{g'(x_K)}\frac{1}{v_K} - x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\big(- \frac{\partial S^*}{\partial \sigma_\theta} + \frac{S^*}{\sigma_\theta}\big) \\
&= -\frac{\partial S^*}{\partial \sigma_\theta} - x_K + \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{\partial S^*}{\partial \sigma_\theta} - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{S^*}{\sigma_\theta}
\end{aligned}$$

Because we know that the teacher wants to set S^* such that $f(S^*) = 0$

$$\begin{aligned}
0 &= -\frac{\partial S^*}{\partial \sigma_\theta} - x_K + \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{\partial S^*}{\partial \sigma_\theta} - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{S^*}{\sigma_\theta} \\
\frac{\partial S^*}{\partial \sigma_\theta} &= \frac{-x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{S^*}{\sigma_\theta}}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})}
\end{aligned}$$

I need to show that the denominator is positive, so the whole expression is negative. This is equivalent to showing that $S^* > \sigma_\theta x_K$:

$$\begin{aligned}
1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta}) &> 0 \\
g(\frac{-S^*}{\sigma_\theta}) &< \frac{\sigma_\theta}{v_K} \\
\frac{-S^*}{\sigma_\theta} &< g^{-1}(\frac{\sigma_\theta}{v_K}) \\
\frac{-S^*}{\sigma_\theta} &< x_K \\
S^* &> \sigma_\theta x_K
\end{aligned}$$

To show that $S^* > \sigma_\theta x_K$, I first take derivatives of f with respect to S^*

$$\begin{aligned}
\frac{\partial f}{\partial S^*} \quad & -1 + g(\frac{-S^*}{\sigma_\theta})\frac{v_K}{\sigma_\theta} \gtrless 0 \\
& g(\frac{-S^*}{\sigma_\theta})\frac{v_K}{\sigma_\theta} \gtrless 1 \\
& g(\frac{-S^*}{\sigma_\theta}) \gtrless \frac{\sigma_\theta}{v_K} \\
& -\frac{S^*}{\sigma_\theta} \gtrless g^{-1}(\frac{\sigma_\theta}{v_K}) \\
& -S^* \gtrless \sigma_\theta x_K \\
& S^* \lesseqgtr \sigma_\theta x_K
\end{aligned}$$

Thus, f is increasing in S^* for $S^* \leq \sigma_\theta x_K$. Finally, I need to show that the S^* such that $f = 0$ is $S^* > \sigma_\theta x_K$. Since f is decreasing in S^* for $S^* > \sigma_\theta x_K$, I only need to show that $f(\sigma_\theta x_K) > 0$ and $f(0) > 0$

$$f(\sigma_\theta x_K) = v_K G(x_K) - \sigma_\theta x_K - \sigma_\theta x_K - v_K G(-x_K) = 2[v_K G(x_K) - \sigma_\theta x_K] - v_K$$

$$f(0) = v_K G(x_K) - \sigma_\theta x_K - 0x_K - v_K G(0) = v_K G(x_K) - \sigma_\theta x_K - v_K/2 = f(\sigma_\theta x_K)/2$$

Because f is decreasing in S^* for $S^* > \sigma_\theta x_K$ and increasing in S^* for $S^* \in [0, \sigma_\theta x_K]$, $S^* = \sigma_\theta x_K$ is a global maximum. If an interior solution exists, then $f(\sigma_\theta x_K) > 0$ and hence $f(0) > 0$. ∎

Proof of Result 1.3

$$\begin{aligned}
\frac{\partial e_K}{\partial \sigma_\theta} &= \frac{\partial S^*}{\partial \sigma_\theta} + x_K + \sigma_\theta \frac{\partial x_K}{\partial \sigma_\theta} \\
&= \frac{-x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{S^*}{\sigma_\theta}}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + x_K + \sigma_\theta \frac{\partial x_K}{\partial \sigma_\theta} \\
&= \frac{-x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{S^*}{\sigma_\theta}}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + x_K + \sigma_\theta \frac{1}{g'(x_K)} \frac{1}{v_K} \\
&= \frac{-x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{S^*}{\sigma_\theta} + x_K - x_K \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + \sigma_\theta \frac{1}{g'(x_K)} \frac{1}{v_K} \\
&= \frac{-\frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})\frac{S^*}{\sigma_\theta} - x_K \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + \sigma_\theta \frac{1}{g'(x_K)} \frac{1}{v_K}
\end{aligned}$$

Because the denominator is positive, the expression is negative. ∎

Proof of Result 1.4

$$\begin{aligned}
\frac{\partial e_K}{\partial \sigma_\theta} &= \frac{\partial S^*}{\partial \sigma_\theta} + x_i + \sigma_\theta \frac{\partial x_i}{\partial \sigma_\theta} \\
&= \frac{-x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta}) \frac{S^*}{\sigma_\theta}}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + x_i + \sigma_\theta \frac{\partial x_i}{\partial \sigma_\theta} \\
&= \frac{-x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta}) \frac{S^*}{\sigma_\theta}}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + x_i + \sigma_\theta \frac{1}{g'(x_K)} \frac{1}{v_K} \\
&= \frac{-x_K - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta}) \frac{S^*}{\sigma_\theta} + x_i - x_i \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + \sigma_\theta \frac{1}{g'(x_K)} \frac{1}{v_K} \\
&= \frac{(x_i - x_K) - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta}) \big(\frac{S^*}{\sigma_\theta} + x_i\big)}{1 - \frac{v_K}{\sigma_\theta} g(\frac{-S^*}{\sigma_\theta})} + \sigma_\theta \frac{1}{g'(x_K)} \frac{1}{v_K}
\end{aligned}$$

When $v_i - v_K$ is large enough, then the first term is positive and larger than the second term. ∎

Proof of Result 1.5

I need to show that $W^A > W^R$ for both the perfect and imperfect information case.

For the perfect information case, I first need to show that the probability that student K+1 passes the class conditional on his level of effort and the level of effort of the other N-1 students in a Nash equilibrium is the following

$$P_{K+1} = K - \sum_{i=1}^{K} P_i$$

Let q_i^s be the probability that student i looses, given that all students choose $e_j \geq v_s^l$, where v_s^l is the lower support of student s equilibrium mixed strategy. Then student i's probability of passing the class when K passes are given is $P_i = 1 - q_i^{K+1}$. Also, $q_i^i = (1 - q_1^i - ... - q_{i-1}^i) \frac{1}{K+2-i}$ Then,

$$\begin{aligned}
P_{K+1} &= 1 - q_{K+1}^{K+1} \\
&= 1 - [1 - q_1^{K+1} - ... - q_K^{K+1}] \frac{1}{K + 2 - (K+1)} \\
&= 1 - [(1 - q_1^{K+1}) - (1 - q_2^{K+1}) ... + (1 - q_K^{K+1}) - (K-1)] \\
&= K - [(1 - q_1^{K+1}) - (1 - q_2^{K+1}) ... + (1 - q_K^{K+1})] \\
&= K - [P_1 + P_2 + ... + P_K] \\
&= K - \sum_{i=1}^{K} P_i
\end{aligned}$$

I can use the above in the definition of W^R:

$$\begin{aligned} W^R = \sum_{s=1}^{N} \mathbb{E}[e_s] &= P_{K+1}v_{K+1} + \sum_{i=1}^{K}(P_i v_i - v_i + v_{K+1}) \\ &= (K - \sum_{i=1}^{K} P_i)v_{K+1} + \sum_{i=1}^{K}(P_i v_i - v_i + v_{K+1}) \\ &= K v_{K+1} + \sum_{i=1}^{K}(P_i v_i - v_i + v_{K+1} - P_i v_{K+1}) \\ &= K v_{K+1} - \sum_{i=1}^{K}(v_i - v_{K+1})(1 - P_i) \\ &< K v_{K+1} \\ &< K v_K \\ &= W^A \end{aligned}$$

When there is imperfect information, the ex ante total level of effort is $W^A = K\int_0^{\bar{v}} v_K dF_K(v_K)$, which is higher than the ex ante total level of effort with relative standard, $W^R = K\int_0^{\bar{v}} v_{K+1} dF_{K+1}(v_{K+1})$. ∎

Proof of Result 1.7 With perfect information, students with high ability exert effort $\mathbb{E}[e_i] < v_{K+1} \leq v_i$ under the relative standard, which is lower than the amount of effort under the absolute standard, $e(v_i) = S^* = v_K$. Students with low ability exert zero effort under both schemes, with the exception of the student with ability v_{K+1}, who exerts zero effort when the standard is absolute but exerts effort $\mathbb{E}[e_i] = P_{K+1}v_{K+1}$ with the relative standard[9].

Under imperfect information, nondecreasing strategies imply that all students exert effort greater than or equal to zero under the relative standard. For students with $v_i < S^*$, their effort level is thus (weakly) lower when the standard is absolute. This and Result 1.7 prove that the opposite is true for students with $v_i > S^*$. ∎

[9]It can be shown that $P_{K+1} > 0$. Suppose not. Then the K students with highest v_i will pass the class with probability 1, so they will exert $\varepsilon > 0$ effort, with ε arbitrary low. In that case, it is not optimal for students with lower v_i to exert zero effort.