INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

578
2024

Teacher Value-Added and the Test Score Gender Gap

A. García-Echalar, S. Poblete y T. Rau.

www.economia.uc.cl · ISSN (edición impresa) 0716-7334 · ISSN (edición electrónica) 0717-7593

Teacher Value-Added and the Test Score Gender Gap*

Andrés García-Echalar   Sebastián Poblete   Tomás Rau

June 9, 2024

Abstract

This paper assesses the effect of teachers on the gender gap in student test scores. It combines different empirical strategies from the value-added and labor economics literature to estimate teacher value-added and its contribution to the math and reading gender gaps. We use rich administrative data from Chile, that allows us to follow teachers through different classes in different years. Our main findings indicate that teachers explain up to 18% of student test score variance and help reduce the gender gap in math by 16.9%. The reduction in the math gender gap is greater in voucher schools (16.1%), among students with more educated mothers (24%) and among those with female math teachers (32.2%). We provide evidence supporting a within-class effect instead of sorting (between-class effect). We conduct several tests and robustness checks to assess the reliability of our findings.

*Corresponding author: Tomás Rau, Instituto de Economía, Pontificia Universidad Católica de Chile and IZA; email, trau@uc.cl. Andrés García-Echalar, Universidad de los Andes, Chile; email, agarcia@uandes.cl. Sebastián Poblete, Department of Economics, Northwestern University; email, sebapoblete@u.northwestern.edu. We would like to thank the seminar participants at SECHI 2022, LACEA/LAMES 2022 and EALE 2023 for their valuable comments. We thank the Education Quality Agency for granting access to the data used in this research. All the results are solely our responsibility and do not represent the stance of the aforementioned institutions. Tomás Rau is grateful for the funding granted by Fondecyt, Project No. 1230984 and the ANID-Millennium Science Initiative Program, Code: NCS2021072. Andrés García-Echalar is grateful for the funding granted by Fondecyt, Project No. 11220294. The usual disclaimers apply.

1 Introduction

Gender differentials in educational outcomes have been extensively studied in the empirical literature, documenting a significant gender gap in standardized math test scores. This gap is present in developed (Mead, 2006; Guiso et al., 2008) and low- and middle-income (Bharadwaj et al., 2016; Guiso et al., 2008) countries. It appears in the early years of schooling and then broadens with age (Fryer and Levitt, 2010). The importance of studying this gap lies in the fact that math test scores are a strong predictor of wage levels (Paglin and Rufolo, 1990; Altonji and Blank, 1999; Murnane et al., 2000, 1995; Weinberger, 1999, 2001) and career choices (see Altonji et al. (2012) for a review). In particular, women are underrepresented in science, technology, engineering, and mathematics (STEM) fields. This gender disparity has implications not only for the overall productivity of the scientific field but also for women who aspire to enter STEM and find themselves at a disadvantage due to underperformance in standardized math testing. This circumstance effectively denies them the opportunity to pursue STEM careers, which often offer elevated income prospects and employment stability (Weinberger, 2001).

A growing body of literature documents differences in teacher effectiveness by student characteristics, such as gender, race, and socioeconomic status (Fox, 2016; Bates et al., 2022; Aucejo et al., 2022; Bryan S. Graham and Zamarro, 2023; Biasi et al., 2021). However, as noted by Delgado (2023), it is unclear whether these differences are actually caused by the teachers themselves or if they occur because of the sorting of students to teachers. Understanding whether these variations are due to the real impact of teachers or just to the way in which students are sorted is important for policymaking. If some teachers are particularly good at teaching certain groups of students, reallocating teachers could improve the achievements of all students and reduce learning gaps.

In this paper, we contribute to the literature by estimating the impact of teachers on reading and math gender gaps. We use administrative data from Chile that match student characteristics and outcomes to teachers, allowing us to identify and compute teacher value-added (VA) in student test scores. We examine whether teachers differentially provide VA to female and male students by implementing the Chetty et al. (2014) VA estimator separately for each gender. This approach yields two distinct VA measures for each teacher: one that represents the average impact of the teacher on male students' scores and another for female students' scores.

The Chilean case helps shed light on the math gender gap issue for two reasons. First, Chile exhibits significant gender differences in school and the labor market despite advances in the last decades. Second, the availability of rich administrative test score data at the student level from Chile's *Ministry of Education* permits matching students' results with teacher information via unique identifiers. These facts render the Chilean case informative regarding the math gender gap.

The estimation of any teacher VA model faces the identification challenge of addressing potential student–teacher sorting and ensuring that observed differences in student outcomes genuinely reflect the causal effect of teachers. As shown by Chetty et al. (2014), including a student's own lagged test scores as control variables makes the forecast bias of the estimate negligible and statistically insignificant. To mitigate this identification concern, we incorporate controls for students' prior test scores and socioeconomic characteristics when estimating our VA models.

To further reduce identification concerns, we provide quasi-experimental evidence based on teacher staff changes that rules out the presence of forecasting bias in our teacher VA estimates. We exploit naturally occurring teacher turnover as in Chetty et al. (2014), enhancing the validity of our findings on the impact of teachers on student outcomes.

Once we estimate the teacher VA for boys and girls, we compute the contribution to the test score gender gap for both math and reading and apply Card et al. (2016) (CCK hereafter) to decompose the teacher VA contribution to the gender gap into teacher–student sorting (between-class) and a within-class effect. As discussed in the following section, the sorting effect manifests as gender test score differences due to the different distributions of boys and girls across classrooms. Hence, it measures the effect of between-classroom gender segregation on the test score gap that appears, for example, if boys are overrepresented in classes with teachers with higher VA and vice versa. The presence of this channel has been documented in Italy by Carlana (2019). On the other hand, the within-class effect refers to the portion of the gender test score gap that results from interactions within a classroom after removing the effect of gender segregation between classrooms. The mechanisms behind this "within-class" effect may be due to teacher or student behavior, with the former referred to as "teacher bias" (Sansone, 2017; Alan et al., 2018; Carlana, 2019) and the latter as "role model" (Dee, 2007; Hoffmann and Oreopoulos, 2009; Carrell et al., 2010).

For reading, we find that the male-female gender gap (-0.211 SD) is explained in a small fraction (7.1%) by teacher VA; a similar contribution is found among male and female teachers when analyzing the effects in public and voucher schools. On the other hand, in math, where boys outperform girls by 0.071 SD, we observe that teacher VA is greater for girls than for boys. This difference in value-added between genders is -0.012 SD, meaning that teacher VA contributes to reducing the gap by 0.012 SD. This difference accounts for 16.9% of the overall math gap.

Regarding teacher gender, the math gender gap is smaller for students with female teachers (0.059 SD) than for those with male teachers (0.095 SD). While teacher VA for boys and girls does not differ for male teachers, female teachers' VA is greater for girls than for boys. This teacher VA difference accounts for 32.2% of the overall math gap. Interestingly, approximately 3/4 of this effect is due to the within-class effect, and the rest is due to the sorting effect.

When we conduct an analysis by school type, we find that the math gender gap is larger in voucher schools than in public schools (0.09 SD vs. 0.06 SD), where the differences in teacher VA between boys and girls are -0.007 SD and -0.015 SD, accounting for 10.9% and 16.1% of the total gap, respectively. In both cases, the channel is a within-class effect. Finally, the math gender gap is lower the higher the level of mother's education. For students with mothers with tertiary education the differences in teacher VA between boys and girls is -0.014 SD, accounting for 24% of the total gap.

While many studies have focused on the effect of teacher characteristics on student achievement (Dee, 2007; Hoffmann and Oreopoulos, 2009; Carrell et al., 2010; Paredes, 2014), we focus on gender differences in teacher effectiveness across student genders, expanding the approach of Chetty et al. (2014) and implementing nonparametric decompositions from Card et al. (2016). This empirical approach allows us to assess the direct effect of teacher VA on the student test score gap and to disentangle whether this effect is due to a sorting channel or a within-class effect.

Our paper relates to the literature stream on disparities in teacher impacts by student type. In particular, our paper relates to the work of Aucejo et al. (2022), Delgado (2023) and Barrios-Fernandez and Riudavets-Barcons (2024), who study the gender differential effects of teachers on different types of student test score gaps. They implement a similar approach to ours by estimating a student type-specific teacher VA measure (student race and gender) using the estimator proposed in Chetty et al. (2014) broken out by student type. Our paper complements theirs by implementing a between- and within-class decomposition (CCK) to these student-type-specific teacher VA effects on the gender gap in student test scores.

The rest of the article is organized as follows. Section 2 describes the methods used for estimating teacher VA and decomposing the contribution to gender gaps into sorting and within-class effects. Section 3 details the institutional background and data, while Section 4 discusses the results. Section 5 concludes the paper.

2 Methods

2.1 Teacher value-added model

To assess the effect of teachers on student test scores, we follow Chetty et al. (2014). Recent literature has mentioned their approach as one of the most convincing in reducing forecast bias, even in the presence of sorting (Kane and Staiger, 2008; Delgado, 2023; Aucejo et al., 2022).

To allow for differentiated effects of teachers for girls and boys, we estimate teacher VA measures separately by student gender, exploiting matched student–teacher panel data. Hence, in a given year t, a student i of gender $G(i) \in \{\text{male, female}\}$ is assigned to a classroom $c = c(i,t)$ and is taught by a teacher $j = j(c)$. The student's test score is denoted by A^*_{it}, where $\mu^{G(i)}_{jt}$ is teacher VA in year t for students of gender $G(i)$. Note that we allow teacher VA terms $\mu^{G(i)}_{jt}$ to vary by student gender. The empirical model is as follows:

$$
\begin{aligned}
A^*_{it} &= \boldsymbol{X}_{it}\beta^{G(i)} + \nu_{it}, \\
\nu_{it} &= \mu^{G(i)}_{jt} + \theta_c + \tilde{\varepsilon}_{it}
\end{aligned}
\tag{1}
$$

where $\boldsymbol{X}_{it}$ are observable determinants of test scores, such as lagged test scores, student and classroom characteristics that vary over time.[1] We decompose the error term ν_{it} into the gender-specific teacher VA $\mu^{G(i)}_{jt}$, exogenous classroom shock θ_c and idiosyncratic exogenous shock to test scores $\tilde{\varepsilon}_{it}$. Let $\varepsilon_{it} \equiv \theta_c + \tilde{\varepsilon}_{it}$ be the unobserved component in student achievement that is unrelated to teacher quality.

While there are no restrictions on how $\mu^{G(i)}_{jt}$ can vary over time, we need to assume that teacher VA and student test scores follow a covariance stationary process. Hence, $E[\mu^{G(i)}_{jt}|t] =$

[1] We exclude parent characteristics, such as household income and education level, from our control variables $\boldsymbol{X}_{it}$ to test for selection on observables, as in Chetty et al. (2014). More details are provided in Section 4.2.

$E[\varepsilon_{it}|t] = 0$; $cov(\mu_{jt}^{G(i)}, \mu_{j,t+s}^{G(i)}) = \sigma_{\mu s}$ and $cov(\varepsilon_{it}, \varepsilon_{i,t+s}) = \sigma_{\varepsilon s}$ for all t.[2]

The key empirical challenge is that the characteristics $\boldsymbol{X}_{it}$ can be correlated with ε_{it}; therefore, we must account for that selection to estimate $\mu_{jt}^{G(i)}$. For example, if students who achieve high test scores are tracked by the principal and assigned to better teachers, then our estimates of teacher fixed effects would be upward biased. The intuition and main assumption behind the estimator described below is that controls $\boldsymbol{X}_{it}$ are sufficiently rich that any remaining unobserved heterogeneity is balanced (on average) across teachers. Following our example, high-achieving students who are assigned to better teachers probably received high test scores in the previous period; therefore, conditional on past test scores (which we control for), the remaining variation in test scores should be due to the teachers themselves and conditionally uncorrelated with unobservables ε_{it}, which we assume are conditionally balanced across teachers.

As Chetty et al. (2014), we estimate teacher VA in three steps. First, we estimate $\beta^{G(i)}$ by exploiting within-teacher variation using the following regression function:

$$A_{it}^{*} = \alpha_j + \boldsymbol{X}_{it}\beta^{G(i)}, \tag{2}$$

where α_j is a teacher fixed effect. We then compute the residualized test scores:

$$A_{it} = A_{it}^{*} - \boldsymbol{X}_{it}\beta^{G(i)} = \mu_{jt}^{G(i)} + \varepsilon_{it} \tag{3}$$

Finally, using the residualized test scores A_{it}, we compute $\bar{A}_{jt}$, which represents the average residual test scores of all the students of teacher j in period t. Then, we obtain our teacher

[2]As noted by Chetty et al. (2014), this assumption simplifies the estimation of teacher VA by decreasing the number of parameters to be estimated.

VA measure as the best linear predictor of $\bar{A}_{jt}$ based on past scores:

$$\hat{\mu}^G_{jt} = \sum_{s=1}^{t-1} \hat{\psi}_s \bar{A}_{js} \tag{4}$$

The vector $(\hat{\psi}_1, ..., \hat{\psi}_{t-1})$ is chosen such that it minimizes the (mean-squared) forecast error of the average residualized test scores:

$$(\hat{\psi}_1, ..., \hat{\psi}_{t-1}) = \underset{(\psi_1,...,\psi_{t-1})}{\text{argmin}} \sum_j \left(\bar{A}_{jt} - \sum_{s=1}^{t-1} \psi_s \bar{A}_{js} \right)^2 \tag{5}$$

The vector $(\hat{\psi}_1, ..., \hat{\psi}_{t-1})$ corresponds to shrinkage factors for our VA estimates to reduce the incidence of small sample bias. Hence, the value-added estimate is weighted by a shrinkage coefficient to correct the possible temporal variation of teacher quality.

Note that our VA estimate in Equation 4 is a leave-year-out measure of teacher VA. As noted by Chetty et al. (2014), if $\bar{A}_{jt}$ is not left out, any regression of student test scores in year t on teacher VA would include the same estimation errors on the left- and right-hand sides of the regression. This would lead to biased estimations of the true teachers' causal effects.

2.2 Normalization of teacher VA estimates

Given that teacher VA is a zero-mean measure, to assess the average contribution of teachers to test the score gender gap in the entire sample $(E[\mu^M_{jt}] - E[\mu^F_{jt}])$, we need to renormalize the VA estimates relative to a reference group. One plausible group is teachers with low scores on the College Entrance Exam (CEE), since it has been documented that there is a strong positive correlation between teachers' precollege academic achievement and several measures of teacher productivity (Gallegos et al., 2019).[3]

[3]The College Entrance Exam is mandatory for most applicants to Chilean universities and has been implemented since 1967. The CEE includes two mandatory tests, reading and math, and up to three optional tests, science (biology, physics, and chemistry), history and social science.

Hence, we normalize the μ^G_{jt} estimates by setting teacher VA (by student gender, subject and year) for a group of "low-CEE" teachers to zero. A simple way of implementing this is by defining "low-CEE" teachers as those in the tenth percentile of the CEE distribution.

2.3 Decomposing the effect of teacher VA

To analyze the impact of teacher VA on math and reading gender gaps, we follow the approach of Card et al. (2016), which is based on an Oaxaca-style decomposition into teacher–student sorting (between-class) and within-class effects (Oaxaca, 1973; Fortin et al., 2011).

For simplicity, we use *male* and *female* as shorthand for the respective events $G(i) = male$ and $G(i) = female$. Then, using Equation 1, the gender gap in test scores is given by:

$$\begin{aligned} E\left[A^*_{it}|male\right] - E\left[A^*_{it}|female\right] = {} & E\left[\boldsymbol{X}_{it}\beta^M + \mu^M_{jt} + \theta_c + \tilde{\varepsilon}_{it}|male\right] \\ & - E\left[\boldsymbol{X}_{it}\beta^F + \mu^F_{jt} + \theta_c + \tilde{\varepsilon}_{it}|female\right] \end{aligned} \tag{6}$$

where $E\left[\mu^M_{jt}|male\right]$ and $E\left[\mu^F_{jt}|female\right]$ correspond to the average VA received by boys and girls, respectively. The difference between these two terms is the contribution of teachers to the gender gap, and we decompose the difference into sorting and within-class effects by:

$$\begin{aligned} E\left[\mu^M_{jt}|male\right] - E\left[\mu^F_{jt}|female\right] = {} & E\left[\mu^M_{jt} - \mu^F_{jt}|male\right] \\ & + E\left[\mu^F_{jt}|male\right] - E\left[\mu^F_{jt}|female\right] \end{aligned} \tag{7}$$

$$\begin{aligned} = {} & E\left[\mu^M_{jt} - \mu^F_{jt}|female\right] \\ & + E\left[\mu^M_{jt}|male\right] - E\left[\mu^M_{jt}|female\right] \end{aligned} \tag{8}$$

Similar to Card et al. (2016), the first right-hand side term in Equation 7 represents the average within-class effect, calculated by comparing VA measures μ^M_{jt} and μ^F_{jt} across the distribution of boys. The second line of Equation 7 is the average sorting effect that results from comparing the average VA for female students, μ_{jt}^F, across the distributions of boys and girls. Equation 8 presents an alternative decomposition in which the within-class effect is calculated using the distribution of girls, and the sorting effect results from the comparison of VA measures for male students across boys and girls' distributions.

Note that if all teachers face the same split of boys and girls (50/50) in their classrooms, then the sorting component in the CCK decomposition would be zero by construction since $E\left[\mu_{jt}^M|male\right] = E\left[\mu_{jt}^M|female\right]$. Hence, for the sorting component to matter, there needs to be some student gender segregation across classrooms/teachers or, in other words, some variation in the female share across classrooms/teachers. In Section 3, we analyze the distribution of female students across classrooms, showing that variation across classrooms motivates the analysis of the sorting component in the decomposition.

The decomposition of the teacher VA differential is highly relevant since it allows us to identify how these two complementary channels explain gender gaps. On the one hand, a positive (negative) sorting effect arises when female students are less (more) likely to be assigned to a teacher with higher VA. On the other hand, a positive (negative) within-class effect arises when teachers, on average, are more (less) likely to deliver higher VA to male students than female students.

Finally, as discussed by Card et al. (2016), the normalization of the teacher VA will not affect the sorting (between-class) component in the CCK decompositions since the constants will cancel out. However, within-class sex-specific terms might be affected by normalization.

3 Background and Data

The primary educational system in Chile consists of public and voucher schools, which enroll approximately 93% of students, and private schools, enrolling the remaining 7%. For these schools, the *System for Measuring the Quality of Education* (SIMCE) is a standardized instrument for measuring school performance. Every year, all students of certain grades take the test on different grade-specific subjects, such as math and reading. These tests, based on item response theory, such as the TIMSS or PISA tests, reveal a significant math gender gap favoring males.[4] The difference in math scores ranges from 0.17 to 0.19 standard deviations (SD) across several cohorts (Paredes, 2014; Bharadwaj et al., 2016). As in other countries, the difference appears in the early stages of schooling and widens with age (Bharadwaj et al., 2016). Consequently, while only 21% of Chilean STEM graduates are female, the OECD average female representation is 39% (Cruz and Rau, 2022).

Our census data come from several sources. The core data for all 6th-grade students were obtained from records of achievement between 2013 and 2015; this information is publicly available through the *Ministry of Education* (MINEDUC, 2023). With these records, we have student-level information such as gender, age, attendance rate, an indicator for repeating the grade, class size, class female proportion, class identifier, and GPA in 4th and 5th grades. In addition, we use data from the SIMCE, which is available upon request (SIMCE, 2023). For the same period, we have access to standardized test scores in math and reading for all 6th graders along with socioeconomic information such as family income and educational level for both parents. Moreover, we also have access to math and reading test scores two years earlier when students were in 4th grade to control for prior test scores in our VA models, as suggested by Chetty et al. (2014).[5] We link all this information through a student-level identifier.

Regarding teacher-related data, our primary data source consists of records documenting teachers assigned to specific classes obtained from MINEDUC. This dataset encompasses all 6th-grade math and reading teachers and allows us to build a panel dataset covering 2013 to 2015. Each teacher's entry in the dataset contains gender information, and for each academic

[4]For more details about Chile's educational system, see Contreras and Rau (2012) and Rau et al. (2013).

[5]This is why we chose 6th graders as our population of interest. Beginning in 2013, the SIMCE for both math and reading was applied yearly in 6th grade, and the closest evenly spaced available prior test scores were those in fourth grade for the same cohorts.

year, we have individual information, including class identifiers and school characteristics that indicate whether the school is situated in a rural or urban setting and whether it is publicly funded or a voucher school.

Furthermore, we link this dataset with information concerning school average monthly fees, sourced from MINEDUC for all voucher schools. Public elementary schools do not entail any costs.[6] All these data are linked with the student data through the class identifier.[7] Finally, from the *Department of Educational Evaluation, Measurement and Registration* (DEMRE), we obtained CCE scores for approximately 18% of our teacher observations (DEMRE, 2023). We average the math and reading scores and standardize it to build the normalization constants as discussed in Section 2.2.

Table 1 provides summary statistics for all variables at the teacher and student levels annually from 2013 to 2015. We have data from 19,239 teachers in math and reading. Note that 10,411 (54%) of these teachers appeared only once during 2013–2015, whereas 8,828 appeared multiple times. Furthermore, teachers may be assigned to more than one class, with 76% teaching a single class per year and 24% managing two to five classes annually. Approximately 74% of math and reading teachers are female. Over 50% of teachers work at voucher schools in classrooms of about 30 students, where female students represent approximately 48% with an SD of nearly 15%.[8]

[6] We exclude private school teachers (7.46% of the main sample) since there is no public information about fees for these schools. As discussed by Urrea (2018), school fees are essential for the unbiasedness of teacher VA estimates in the Chilean context.

[7] As in Chetty et al. (2014), we restrict our data to the sample of observations that (i) have prior test scores available, (ii) exclude special education, education at home and others, (iii) drop classrooms with fewer than 10 and more than 50 students, (iv) exclude teachers with more than 200 students in a single grade, (v) have nonmissing data on current or prior test scores in the subject for which the VA model is estimated, and (vi) drop classrooms with fewer than 7 observations with current and prior test scores in math or reading.

[8] Teacher characteristics do not vary economically or statistically by the number of years that teachers appear in our study. See Appendix A for details.

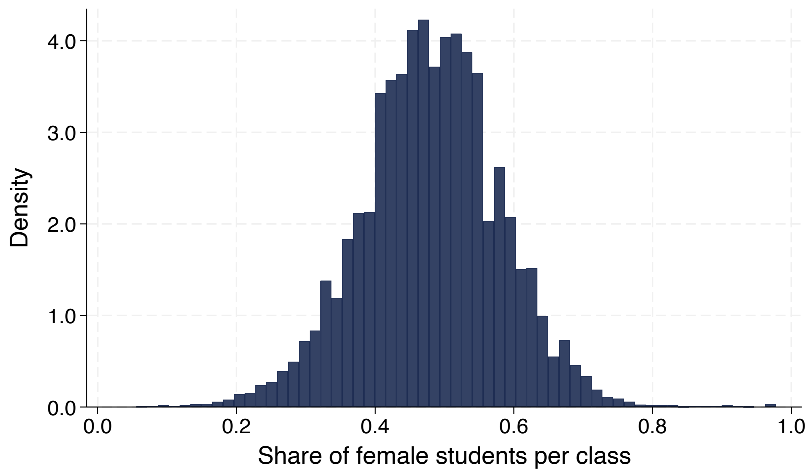
Table 1: Descriptive Statistics

Variable	2013			2014			2015		
	Mean	SD	Obs.	Mean	SD	Obs.	Mean	SD	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Teachers:			**10,639**			**10,723**			**10,465**
Female (%)	74.15	43.78	9,965	74.13	43.79	10,715	75.15	43.22	10,458
Rural School (%)	15.01	35.72	10,639	15.33	36.03	10,723	15.51	36.20	10,465
Voucher School (%)	51.48	49.98	10,639	51.64	49.98	10,721	52.39	49.95	10,463
School Monthly fee (CLP$)	8,651	17,543	9,034	8,862	16,895	9,096	9,409	18,249	8,763
School Imputed fee (%)	56.66	49.56	9,034	56.52	49.58	9,096	56.46	49.58	8,763
Class Size	30.74	9.27	10,639	30.30	9.28	10,723	29.96	9.28	10,465
Class Female Proportion (%)	47.83	14.95	10,639	48.22	14.88	10,723	48.25	14.73	10,465
CCE Score	499.97	80.49	1,691	499.26	79.18	2,113	497.91	78.50	1,841
Students:			**171,586**			**173,369**			**166,573**
Math Score	250.88	47.86	166,959	249.14	48.06	170,152	250.81	48.43	153,621
Reading Score	251.20	48.64	167,011	240.92	50.15	169,677	246.89	52.03	161,865
Female (%)	50.68	50.00	171,586	50.98	50.00	173,369	50.95	50.00	166,573
Age (years)	11.21	0.48	171,585	11.22	0.48	173,368	11.27	0.49	166,573
Attendance (%)	92.86	6.10	171,488	93.02	6.02	173,263	93.03	6.07	166,482
Repeating Grade (%)	0.07	2.67	171,540	0.07	2.58	173,304	0.06	2.55	166,521
GPA in 5th grade	5.67	0.55	171,451	5.72	0.53	173,236	5.74	0.53	166,443
GPA in 4th grade	5.87	0.51	171,477	5.86	0.50	173,202	5.88	0.50	166,503
Father's Education (years)	11.52	3.45	132,434	11.64	3.41	142,119	11.66	3.46	131,447
Mother's Education (years)	11.57	3.30	138,777	11.67	3.26	148,396	11.79	3.30	136,920
Monthly Household Income (CLP$k)	413.86	386.82	138,917	436.74	392.46	149,649	475.10	419.17	137,085

Notes: SD stands for standard deviation. Observations for teachers correspond to 19,239 different people over time. Teachers may be assigned to more than one class and even teach math and reading to the same group of students in a given year.

Finally, regarding the split of male and female students across classrooms, Figure 1 shows the distribution of the share of females across mixed-gender classrooms. There is considerable variation around 0.5, which motivates the sorting component's potential role in CCK decomposition.

Figure 1: Histogram of class female share



4 Results

4.1 Variance decomposition and teacher value-added estimates

We follow the methodology in Chetty et al. (2014) for estimating teacher VA as described in Section 2.1. We begin by estimating test score residuals within each subject (reading and math), for which we regress raw scores (standardized at the subject-year level) on a vector of individual, class, school-year and school covariates along with teacher and year fixed effects.[9]

[9]We replicate the list of covariates used by Chetty et al. (2014) as closely as possible. At the individual level, we include gender, age, attendance rate, an indicator for repeating the grade, cubic polynomials for prior test scores in both math and reading and cubic polynomials for the GPA in 5th and 4th grades standardized at the class level. In addition, we set the other-subject prior score to zero when missing, and we include an indicator variable for other-subject missing data that interacts with the polynomial of prior own-subject scores. At the class level, we add class size, female proportion, and cubic polynomials for average prior scores in both math and reading and averages for age, assistance rate and repeating grade. At the school-year level, we include monthly fee, an indicator for imputed fee and means for female proportion, age, assistance rate and

Second, we estimate the autocovariances and autocorrelations of mean test score residuals across classes taught by a given teacher in different periods, separately for boys and girls and reading and math. Panel A of Table 2 presents the estimation results with findings similar to those of Chetty et al. (2014). Here, correlations represent the reliability of lagged mean test scores by class as predictors of current teacher quality. The greater the autocorrelation, the more effective lagged scores are as predictors of teacher performance. For reading, reliability declines from 0.228–0.261 in the first year to 0.133–0.234 after two years, while it declines from 0.500–0.531 to 0.446–0.460 for math. In addition, we present in Panel B of Table 2 the raw variances of test score residuals by subject and their decomposition into components driven by student-, class-, and teacher-level variation, provided that 6th-grade teachers can teach more than one class per year. Teachers explain only 7.46–7.50% of the variation in reading and 15.53–17.94% in math. In both cases, idiosyncratic student-level variation accounts for approximately 80% of the total raw variation in achievement.

Table 2: Teacher VA Parameter Estimates

	Reading				Math			
	Boys		Girls		Boys		Girls	
Panel A: Autocovariance and Autocorrelation Vectors								
	Covariance	Correlation	Covariance	Correlation	Covariance	Correlation	Covariance	Correlation
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lag 1	0.022	0.228	0.022	0.261	0.048	0.500	0.050	0.531
	(0.002)		(0.001)		(0.002)		(0.002)	
Lag 2	0.011	0.133	0.017	0.234	0.044	0.446	0.045	0.460
	(0.002)		(0.002)		(0.002)		(0.002)	
Panel B: Within-Year Variance Decomposition								
	Variance	%	Variance	%	Variance	%	Variance	%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	0.500	100.00	0.432	100.00	0.399	100.00	0.363	100.00
Individual	0.431	86.13	0.371	85.76	0.323	80.89	0.285	78.52
Class	0.032	6.38	0.029	6.78	0.014	3.58	0.013	3.55
Teacher	0.038	7.50	0.032	7.46	0.062	15.53	0.065	17.94

Notes: Standard errors clustered at the teacher level in parentheses for covariance estimations. Autocovariances and autocorrelations are estimated for mean test score residuals between classes taught by a given teacher, separately for math and reading and boys and girls. Following Chetty et al. (2014), these statistics are measured at one-year and two-year lags.

repeating grade. Finally, at the school level, we add indicator variables for rural/urban and voucher/public and cubic polynomials for average prior scores in both subjects.

Third, we predict each teacher's VA for every year t using test score residuals from all years except year t and the estimated autocovariances presented in Table 2. The SDs of our teacher VA estimates in reading are 0.078 and 0.085 for boys and girls, respectively, while the SDs of teacher VA estimates in math are 0.169 and 0.164 for boys and girls, respectively. These results are also in line with those obtained by Chetty et al. (2014).

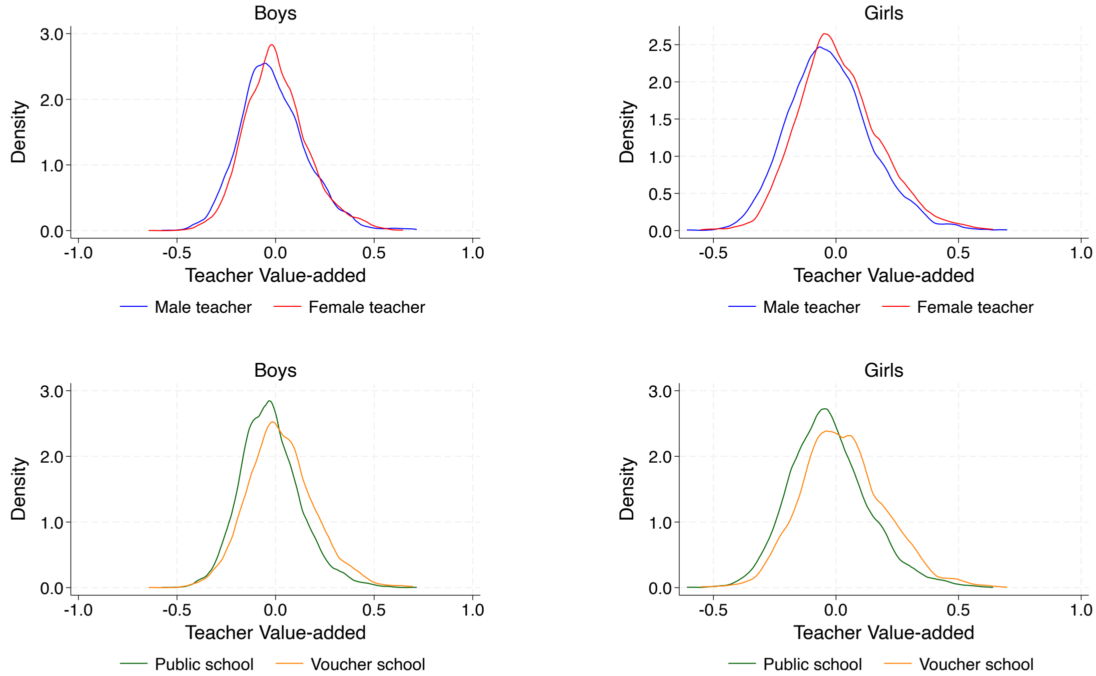
Figure 2 displays the density of the raw teacher VA estimates by teacher/student gender and school type for math. While the distribution of teacher VA for female teachers is clearly to the right of the VA distribution of male teachers when looking at female students, such a pattern is not present for male students. In fact, the mean difference (male–female) in teacher VA is -0.037 SD for girls, while it is -0.018 SD for boys; both are significant at the 1% level. This suggests that girls receive greater VA from female teachers than boys do in math. By type of school (public or voucher), we can see that girls and boys obtain more VA from voucher schools. The mean difference (voucher-public) in teacher VA is 0.058 SD for girls and 0.054 SD for boys, both of which are significant at the 1% level. The corresponding results for reading are presented in Appendix B.

4.2 Reliability and normalization of teacher VA estimates

As discussed in Section 2.1, the estimator for teacher j's VA in year t $\hat{\mu}^G_{jt}$ is the best linear predictor of $\bar{A}_{jt}$ based on prior scores. Therefore, as Chetty et al. (2014) point out, a regression of test score residuals A_{it} on teacher VA $\hat{\mu}^G_{jt}$ should yield a coefficient of 1 by construction (see equations 3 and 4) under stationarity. We confirm this by running an OLS regression of A_{it} on teacher VA $\hat{\mu}^G_{jt}$, including fixed effects by subject and student gender with clustered standard errors at the school cohort level. The point estimate is 0.978, and the 95% confidence interval is $(0.948, 1.009)$.

Panel A in Figure 3 presents the relationship between residuals A_{it} and teacher VA estimates

Figure 2: Empirical distributions of math teacher VA estimates by gender and school type

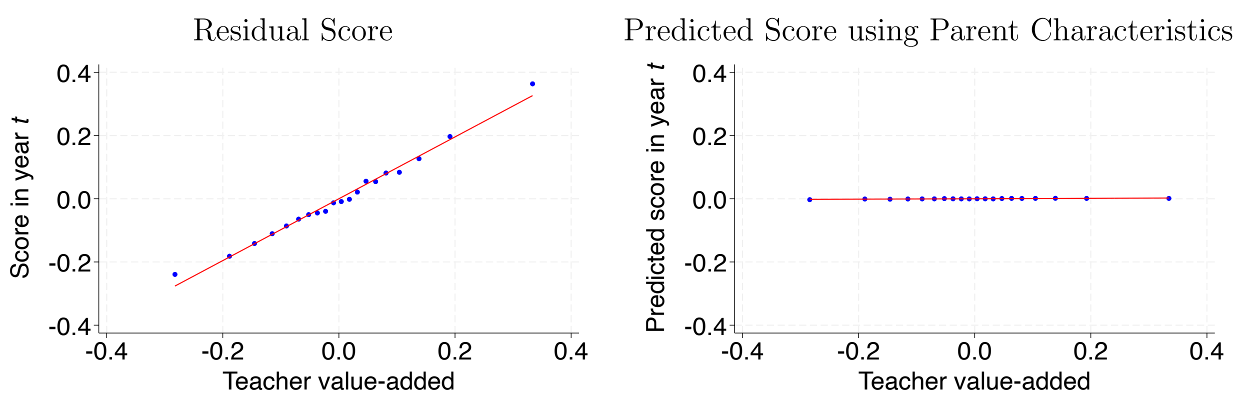


$\hat{\mu}_{jt}^{G}$ nonparametrically. Teacher VA has a 1-1 relationship with test score residuals, showing that the linear prediction model fits the data well.

As Chetty et al. (2014) note, the relationship found between $\hat{\mu}_{jt}^{G}$ and A_{it} could be driven by the true causal effect of teachers μ_{jt}^{G} or by differences in student characteristics ε_{it} that persist across teachers. For this reason, following the authors, we present forecast bias estimates for our model to verify that teacher VA measures are unbiased predictors of teacher quality. First, we focus on the degree of selection on observable characteristics that we excluded from the model to estimate forecast bias. Second, we focus on the bias from selection on unobservables by estimating it through a quasi-experimental approach that exploits teacher turnover.

Bias on observables. We can measure the degree of bias by analyzing whether students are sorted to teachers based on unobserved determinants of scores ε_{it}. Even if we cannot observe ε_{it}, we can nonetheless obtain partial information about ε_{it} by using variables that predict score residuals A_{it} but were omitted from our VA model. These variables are parental characteristics (household income and education level of both parents), and assuming that students are sorted to teachers purely on these variables, we can estimate forecast bias by regressing predicted test scores based on these characteristics that were excluded from the VA model on teacher VA estimates. The degree of forecast bias due to selection on parental characteristics is 0.7% since this regression's coefficient is 0.0067. The 95% confidence interval upper bound for this estimate is 0.8%. Panel B in Figure 3 presents the nonparametric analog of this linear regression, which is nearly flat throughout the distribution, confirming that forecast bias from selection on observable parent characteristics excluded from our VA model is negligible. Furthermore, and as expected, including these characteristics as control variables results in virtually the same VA estimates.

Figure 3: Reliability of Teacher VA Estimates: Effects of VA on Actual and Predicted Scores



Panel A

Panel B

Bias on unobservables. The previous analysis does not rule out the possibility that students are sorted to teachers based on unobservable characteristics orthogonal to parent characteristics. For this reason, we use the proposed quasi-experiment that exploits teacher turnover to assess bias in a complementary way. In our sample, 30% of teachers move to a different grade the following year within the same school, 15% of teachers move to a different school, and another 20% drop from our sample entirely. Therefore, adjacent student cohorts in a given school are usually exposed to different teachers. We use this naturally occurring teacher turnover to estimate forecast bias by comparing the change in mean scores across cohorts to the change in mean VA driven by teacher turnover. As in Chetty et al. (2014), we can identify the degree of bias by assuming that the changes in VA within a school are uncorrelated with changes in other determinants of student achievement.[10]

We begin by constructing leave-two-year-out estimates of teacher VA for each teacher using all years except $t-1$ and t and averaging the VA for boys and girls weighted by the gender proportion within the classroom. Then, we calculate the change in mean teacher VA for each school-subject-year cell that represents the changes in the quality of the teaching staff. Finally, we regress the changes in mean test scores across cohorts on changes in the quality of teaching staff. We find that changes in the quality of teaching staff strongly predict changes in test scores across cohorts in a school-subject cell. The estimated coefficient is 0.836 with a standard error of 0.115, implying a statistically nonsignificant bias of 16.42%.

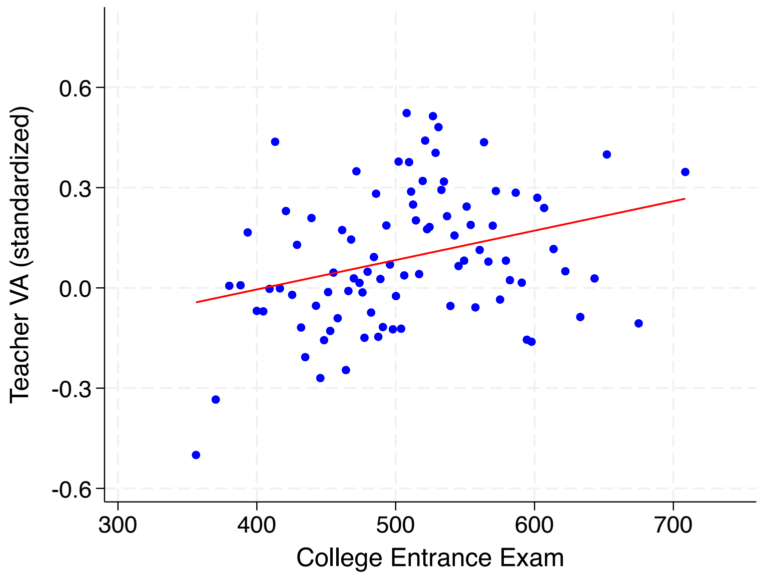
Normalization of teacher VA estimates. As discussed in Section 2.2, we normalize our VA estimates relative to a reference group. In particular, those at the tenth percentile of the teachers' CEE distribution were included. We do so because of a documented positive correlation between teachers' precollege measures of teacher productivity.[11]

Figure 4 shows a bin scatter of teacher VA and teacher CCE. There is a positive correlation between the two measures. Thus, on average, teachers with low CEE scores have low VA. This finding agrees with those of Gallegos et al. (2019) on teacher precollege achievement and short- and long-run measures of teacher productivity.

[10]This quasi-experimental approach exploiting teacher turnover to test for the presence of bias on unobservables has been successfully implemented by other authors in the literature (Adnot et al., 2017; Urrea, 2018; Delgado, 2023).

[11]In Appendix C, we present an alternative normalization of teacher VA using male teachers in rural schools as the reference group and find similar results.

Figure 4: Teachers' Value-Added and College Entrance Exam



To implement the normalization, we use the average score between math and reading as our measure of CEE (the mandatory subjects). Once we estimate the normalization constants, we normalize the teacher VA measures (subtracting those constants) for the entire sample.

Finally, Figure 5 presents the nonparametric relationship between teacher VA and the gender gap attributed to teachers for math and reading in Panels A and B, respectively. For every teacher-year class, an overall VA is calculated as the average between the normalized VA for boys $\widehat{\mu}_{jt}^{M}$ and the corresponding VA for girls $\widehat{\mu}_{jt}^{F}$, weighted by the student gender proportion in class:

$$\widehat{\mu}_{jt} = \widehat{\mu}_{jt}^{M} \cdot (1 - \%\text{Girls}_{jt}) + \widehat{\mu}_{jt}^{F} \cdot \%\text{Girls}_{jt}$$

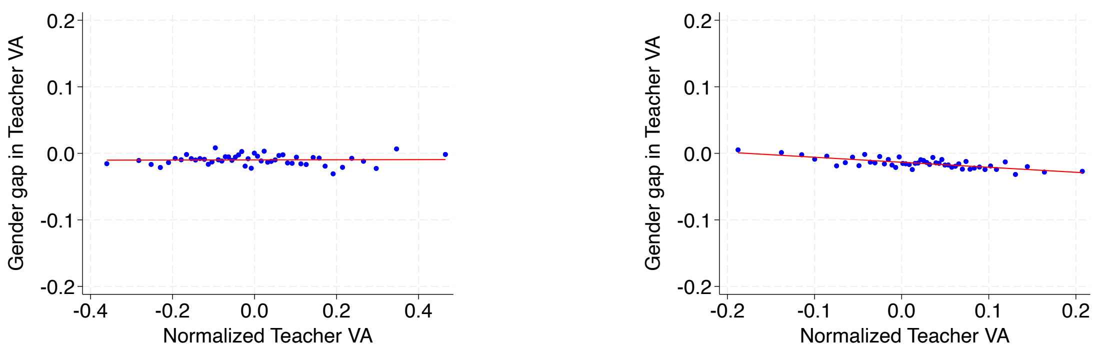
In addition, for every teacher-year class, we calculate the gender gap in VA as the difference between the normalized VA for boys and girls:

$$\text{TVA gap} = \widehat{\mu}_{jt}^{M} - \widehat{\mu}_{jt}^{F}$$

The relationship between the gender gap in teacher VA and average teacher VA is flat through VA support for math teachers (Panel A) and slightly decreasing for reading teachers (Panel B). However, the slope is minimal. Hence, better teachers do not necessarily exhibit larger/smaller gender gaps for boys or girls. We further investigate these results in the next section.

Figure 5: Relationship between Teacher VA and Gender Gap

Math Reading



Panel A Panel B

4.3 CCK decomposition

Table 3 presents the results from the decomposition depicted in Equations (7) and (8). This analysis excludes single-sex schools (approximately 4% of our sample) since we need teacher VA estimates for both male and female students. Column (1) shows the (male–female) gender gap in SDs, and Columns (2) and (3) show the means of teacher VA for each gender. Column (4) reports the total contribution of teachers to the test score gender gap. The sorting (between class) and within-class effects are shown in Columns (5)–(8). Finally, the numbers in parentheses represent the percentage of the test score gap explained by teachers' VA.

In Panel A of Table 3, we can see that the contribution of teacher VA to the gender gap differs by subject. While teacher VA tends to increase the reading gender gap by 7.1% (favoring females), it helps decrease the math gender gap by 16.9% (favoring females). When breaking down the teachers' contribution to the test score gender gap into sorting and within-class effects, we observe that the predominant factor is the within-class effect.

Table 3: Contribution to Gender Gap and CCK Decomposition

					Decompositions of Contribution:			
		Means of TVA		Total	Sorting (between-class)		Gender specific (within-class)	
	Gender Gap	among Boys	among Girls	Contribution of TVA	Using Male Effects	Using Female Effects	Using Male Distribution	Using Female Distribution
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A – Overall Sample:								
Math	0.071	-0.005	0.007	-0.012	-0.002	-0.002	-0.010	-0.010
				(-16.9)	(-2.8)	(-2.8)	(-14.1)	(-14.1)
Reading	-0.211	0.013	0.029	-0.015	-0.000	-0.001	-0.014	-0.015
				(7.1)	(0.0)	(0.5)	(6.6)	(7.1)
Panel B – By Teacher Gender:								
Math								
Male	0.095	-0.013	-0.014	0.001	0.001	0.001	-0.000	-0.001
				(1.1)	(1.1)	(1.1)	(-0.0)	(-1.1)
Female	0.059	-0.000	0.018	-0.019	-0.004	-0.004	-0.015	-0.015
				(-32.2)	(-6.8)	(-6.8)	(-25.4)	(-25.4)
Reading								
Male	-0.210	0.011	0.022	-0.011	-0.002	-0.002	-0.009	-0.009
				(5.2)	(1.0)	(1.0)	(4.3)	(4.3)
Female	-0.211	0.013	0.029	-0.016	0.000	-0.001	-0.015	-0.016
				(7.6)	(-0.0)	(0.5)	(7.1)	(7.6)
Panel C – By School Dependency:								
Math								
Public	0.064	-0.033	-0.027	-0.007	-0.001	-0.002	-0.005	-0.006
				(-10.9)	(-1.2)	(-3.1)	(-7.8)	(-9.4)
Voucher	0.093	0.020	0.035	-0.015	-0.001	-0.001	-0.014	-0.014
				(-16.1)	(-1.1)	(-1.1)	(-15.1)	(-15.1)
Reading								
Public	-0.222	0.016	0.032	-0.016	-0.000	-0.001	-0.015	-0.016
				(7.2)	(0.0)	(0.5)	(6.8)	(7.2)
Voucher	-0.193	0.011	0.026	-0.015	0.000	-0.001	-0.014	-0.015
				(7.8)	(-0.0)	(0.5)	(7.3)	(7.8)

Notes: We restrict the analysis to the sample of teachers in mixed classes where teacher VA is estimated for boys and girls, therefore excluding those teachers in single-sex classes. Figures in parentheses represent the percent of the overall gender gap.

In Panel B, we analyze the heterogeneous effects of teacher VA on the test score gender gap by teacher gender. First, in Column (1), the math gender gap is smaller when students are assigned to female teachers and greater when they are assigned to male teachers. Female teachers reduce the math gender gap by 32.2%, while male professors increase it by 1.1%. Furthermore, the CCK decomposition allows us to better understand the underlying mechanisms involved. As shown by Columns (5) to (8) of Panel B, while for male teachers there is not much variation, approximately three-fourths of female professors' contribution to the math gap is explained by the within-class effect (25.4 pp out of 32.2%), and the other one-fourth is explained by sorting.

Panel C analyzes the heterogeneous effects of teacher VA on the student test score gap by school dependency. For math, while teachers in public schools help reduce the gap by 10.9%, teachers in voucher schools show a stronger effect, decreasing the gap by 16.1%. When analyzing whether the effects are due to sorting or within-class effects, we find that most are due to within-class effects, especially in voucher schools (15.1 pp out of 16.1%).

Our findings align with the literature documenting a student–teacher match component in teacher effectiveness. In particular, our results agree with those of Paredes (2014), who finds that while teacher gender does not correlate with boys' achievement, girls benefit from being assigned to female teachers in Chile. This favorable effect on girls may be due to role model effects, which are compatible with our within-class effects. Additionally, using data from the US, Carrell et al. (2010) finds that while teacher gender has only a limited impact on boys, it has a significant effect on girls' math test scores. They hypothesize that the role model effect may be behind their results but do not test for this channel. In another context, Delgado (2023) also documents disparities in teacher impacts by student type apart from gender. The author finds that some teachers are more successful at improving black students' test scores, while others show greater effectiveness with non-black students.

Our findings contribute to the understanding of the effect of teacher gender on student performance. The higher VA of female teachers to girls than boys is mainly determined within the classroom and not through sorting effects. This result rules out the hypothesis that Chilean girls in the 6th grade self-select into schools with low-VA teachers. This sorting channel has been reported to be relevant in Italy by Carlana (2019), who finds that teacher stereotypes make girls "self-select into less demanding high schools, following the track recommendation of their teachers." However, our findings are not comparable, since Carlana (2019) analyzes 8th graders and their educational track decisions after middle school (academic-oriented, technical, and vocational high school), and our sample includes younger students (6th graders) with no such decisions.

4.4 Heterogeneous effects

In Table 4, we assess the presence of heterogeneous effects in two additional dimensions: mother's education and mother's education combined with teacher gender. We focus on math only since the contribution of teachers to the reading gender gap is small.[12]

In Panel A, we analyze the student test score gap for mothers with different levels of education. The math gender gap decreases as mothers' education increases. Additionally, the greater the mother's education, the greater the effect of teachers in reducing the math gender gap. While girls of mothers with primary education see a 12.6% reduction in the math gender gap due to teachers, those of mothers with tertiary education see a 24.1% reduction in the gender gap associated with teachers.

[12]The reading results can be found in Appendix D.

Table 4: Contribution to math gender gap and CCK decomposition, heterogeneous effects

					Decompositions of Contribution:			
		Means of TVA		Total	Sorting (between-class)		Gender specific (within-class)	
	Gender	among	among	Contribution	Using Male	Using Female	Using Male	Using Female
	Gap	Boys	Girls	of TVA	Effects	Effects	Distribution	Distribution
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A – By mother's education, all teachers:								
Primary	0.087	-0.042	-0.030	-0.011	0.004	-0.005	-0.011	-0.010
				(-12.6)	(4.6)	(-5.8)	(-12.6)	(-11.5)
Secondary	0.079	-0.030	-0.018	-0.012	-0.002	-0.004	-0.008	-0.008
				(-15.2)	(-2.5)	(-5.1)	(-10.1)	(-10.1)
Tertiary	0.058	0.008	0.022	-0.014	-0.003	-0.003	-0.011	-0.012
				(-24.1)	(-5.2)	(-5.2)	(-19.0)	(-20.7)
Panel B – By mother's education, female teacher:								
Primary	0.078	-0.041	-0.018	-0.023	0.002	-0.010	-0.017	-0.016
				(-29.5)	(2.6)	(-12.8)	(-21.8)	(-20.5)
Secondary	0.068	-0.028	-0.009	-0.019	-0.003	-0.006	-0.014	-0.014
				(-27.9)	(-4.4)	(-8.8)	(-20.6)	(-20.6)
Tertiary	0.045	0.012	0.032	-0.021	-0.004	-0.004	-0.016	-0.017
				(-46.7)	(-8.9)	(-8.9)	(-35.6)	(-37.8)

Notes: We restrict the analysis to the sample of teachers in mixed classes where teacher VA is estimated for boys and girls, therefore excluding those teachers in single-sex classes. Figures in parentheses represent the percent of the overall gender gap.

Panel B refines the analysis in Panel A, focusing on female teachers only. We observe the same declining pattern in the math gender gap as the mother's education level increases but with lower levels than those of Panel A. For example, the math gender gap for children of mothers with tertiary education is 0.058 SD overall, but the gap for that same group with a female teacher decreases to 0.045 SD. Interestingly, the contribution of female teachers to the gap increases from 24.1% to 46.7% for that same group.

Hence, daughters of more educated mothers with female teachers have a lower math gender gap, and teachers' contribution to reducing the gap is greater. Identifying whether these differences in teacher effectiveness are due to teacher or student behavior is beyond the scope of this paper but indicates an interesting avenue for future research.

5 Conclusions

In this paper, we use rich administrative data from Chile to estimate the effect of teacher VA on the gender gap in student math (which favors boys) and reading (which favors girls) test scores. Our data allow us to follow teachers over time and through different classes, and we combine different empirical strategies (Chetty et al., 2014; Card et al., 2016) to obtain unbiased estimates of teachers' impact on test scores and their contribution to gender gaps. Overall, we find that teachers account for up to 16% of the score variance and that their value-added helps reduce the gender gap in math and tends to increase the gender gap in reading, favoring girls in both cases. In math, the decrease in the gender gap is nearly one-third when girls are assigned to a female teacher.

When analyzing the channel of these effects, we ruled out the sorting channel: girls are not overrepresented in classrooms with low-VA teachers (between-class effect). This effect occurs within the classroom. This within-class effect is larger when girls are assigned to a female math teacher and even larger when they have mothers with tertiary education.

These teacher-student match effects have policy implications since reallocating some teachers may improve female students' achievement in math, reducing the persistent gender gap in that subject.

References

Adnot, M., Dee, T., Katz, V., and Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in dcps. *Educational Evaluation and Policy Analysis*, 39(1):54–76.

Alan, S., Ertac, S., and Mumcu, I. (2018). Gender Stereotypes in the Classroom and Effects on Achievement. *The Review of Economics and Statistics*, 100(5):876–890.

Altonji, J. G. and Blank, R. (1999). Race and gender in the labor market. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, chapter vol. 3c, pages 3144–3259. Eslevier Science, Amsterdam.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics*, 4(1):185–223.

Aucejo, E. M., Fruehwirth, J. C., Kelly, S., and Mozenter, Z. (2022). Teachers and the gender gap in reading achievement. *Journal of Human Capital*, 16(3):372–403.

Barrios-Fernandez, A. and Riudavets-Barcons, M. (2024). Teacher value-added and gender gaps in educational outcomes. *Economics of Education Review*, 100:102541.

Bates, M. D., Dinerstein, M., Johnston, A. C., and Sorkin, I. (2022). Teacher labor market equilibrium and the distribution of student achievement. Working Paper 29728, National Bureau of Economic Research.

Bharadwaj, P., Giorgi, G. D., Hansen, D., and Neilson, C. A. (2016). The Gender Gap in Mathematics: Evidence from Chile. *Economic Development and Cultural Change*, 65(1):141–166.

Biasi, B., Fu, C., and Stromme, J. (2021). Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage. Working Paper 28530, National Bureau of Economic Research.

Bryan S. Graham, Geert Ridder, P. T. and Zamarro, G. (2023). Teacher-to-classroom assignment and student achievement. *Journal of Business & Economic Statistics*, 41(4):1328–1340.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2):633–686.

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers? Gender Bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Contreras, D. and Rau, T. (2012). Tournament incentives for teachers: Evidence from a scaled-up intervention in chile. *Economic Development and Cultural Change*, 61(1):219 – 246.

Cruz, G. and Rau, T. (2022). The effects of equal pay laws on firm pay premiums: Evidence from chile. *Labour Economics*, 75:102135.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, 42(3).

Delgado, W. (2023). Disparate Teacher Effects, Comparative Advantage, and Match Quality. *EdWorkingPaper*, (23-848).

DEMRE (2023). Base de datos del Departamento de Evaluación, Medición y Registro Educacional, 2001-2009. https://demre.cl. Santiago, Chile.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Chapter 1 - decomposition methods in economics. volume 4 of *Handbook of Labor Economics*, pages 1–102. Elsevier.

Fox, L. (2016). Playing to Teachers? Strengths: Using Multiple Measures of Teacher Effectiveness to Improve Teacher Assignments. *Education Finance and Policy*, 11(1):70–96.

Fryer, R. G. and Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–40.

Gallegos, S., Neilson, C., and Calle, F. (2019). Screening and recruiting talent at teacher colleges using pre-college academic achievement. *Industrial Relations Section Working Paper Series*, 636.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880):1164.

Hoffmann, F. and Oreopoulos, P. (2009). A professor like me. *Journal of Human Resources*, 44(2):479–494.

Kane, T. J. and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Working Paper 14607, National Bureau of Economic Research.

Mead, S. (2006). The truth about boys and girls. Working paper, Education Sector.

MINEDUC (2023). Datos Abiertos del Centro de Estudios Mineduc, 2011-2015. https://datosabiertos.mineduc.cl. Santiago, Chile.

Murnane, R. J., Willett, J. B., Duhaldeborde, Y., and Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(4):547–568.

Murnane, R. J., Willett, J. B., and Levy, F. (1995). The Growing Importance of Cognitive Skills in Wage Determination. *The Review of Economics and Statistics*, 77(2):251–266.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Paglin, M. and Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8(1):123–44.

Paredes, V. (2014). A teacher like me or a student like me? role model versus teacher bias effect. *Economics of Education Review*, 39:38–49.

Rau, T., Rojas, E., and Urzúa, S. (2013). Loans for higher education: Does the dream come true? Working Paper 19138, National Bureau of Economic Research.

Sansone, D. (2017). Why does teacher gender matter? *Economics of Education Review*, 61:9–18.

SIMCE (2023). Base de datos de la Agencia de Calidad de la Educación, 2011-2015. https://www.agenciaeducacion.cl/simce/. Santiago, Chile.

Urrea, I. (2018). Teacher value-added in chile. *Tesis para optar al grado de Magister en Economía, Universidad de Chile.*

Weinberger, C. J. (1999). Mathematical college majors and the gender gap in wages. *Industrial Relations*, 38(July):407–413.

Weinberger, C. J. (2001). Is teaching more girls more math the key to higher wages? In King, M. C., editor, *Squaring Up: Policy Strategies to Raise Women's Incomes in the United States*, chapter 11. University of Michigan Press, Michigan.

Appendix

A Teacher Characteristics by Number of Classes

Table A.1 presents yearly summary statistics for teachers, separated by the number of classes they teach in our sample (one class vs. more than one class). Columns (3), (6) and (9) present difference-in-mean tests. Teachers with one or more than one class are balanced across voucher schools and have classes with the same share of female students. Now, while some other variables have a statistically significant difference, the differences are rather small and economically comparable. For example, while 76% of teachers with one class were female in 2013, for those with more than one class, this percentage is 74%. The two percentage point difference is statistically significant but economically negligible.

Table A.1: Teacher Descriptive Statistics by Number of Classes

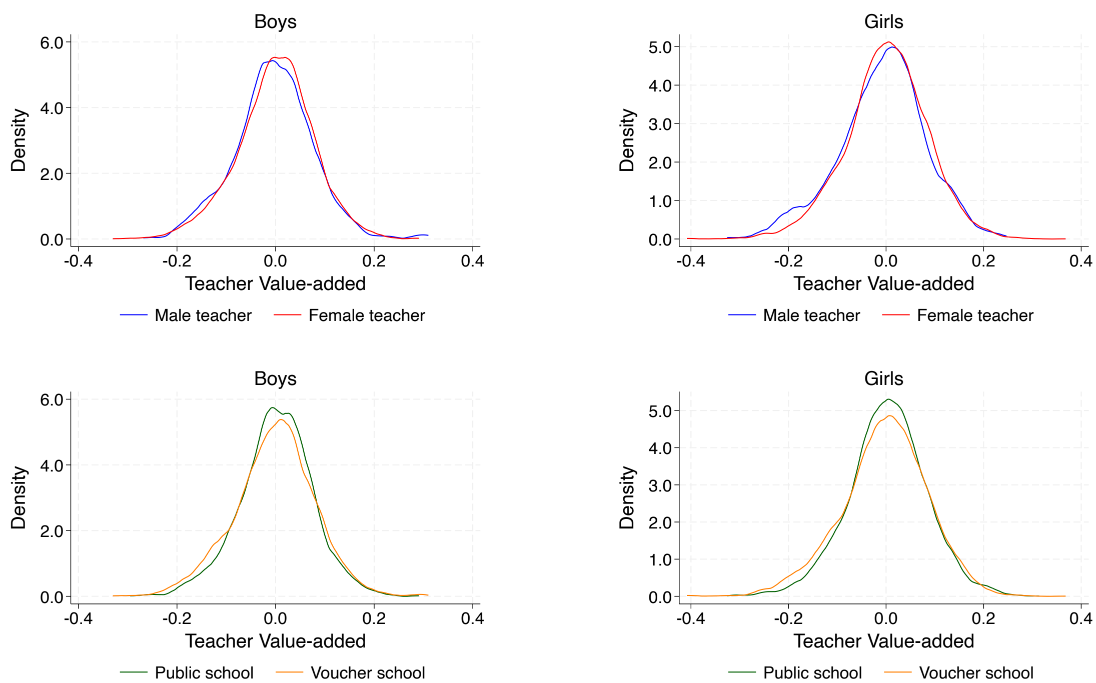
	2013			2014			2015		
Characteristics/# of classes	1	> 1	Diff.	1	> 1	Diff.	1	> 1	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	0.76	0.74	0.02*	0.76	0.74	0.02*	0.77	0.74	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Rural School	0.16	0.15	0.01	0.18	0.14	0.04***	0.17	0.15	0.02*
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
Voucher School	0.51	0.52	-0.01	0.51	0.52	-0.01	0.54	0.51	0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
School Monthly fee (kCLP$)	8.15	8.93	-0.78*	7.92	9.20	-1.28**	9.15	9.53	-0.38
	(0.31)	(0.23)	(0.38)	(0.32)	(0.21)	(0.40)	(0.32)	(0.24)	(0.41)
School Imputed fee	0.58	0.56	0.02	0.59	0.56	0.03*	0.56	0.57	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Class Size	30.07	31.08	-1.01***	29.13	30.73	-1.60***	29.41	30.25	-0.84***
	(0.15)	(0.11)	(0.19)	(0.18)	(0.10)	(0.20)	(0.16)	(0.11)	(0.19)
Class Female Proportion	0.48	0.48	0.00	0.48	0.48	-0.00	0.48	0.48	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations:	**3,853**	**6,786**		**2,922**	**7,801**		**3,636**	**6,829**	

Notes: SE in parenthesis. Diff. presents the difference-in-mean test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Observations for teachers correspond to 19,239 different people over time. Teachers may be assigned to more than one class and even teach math and reading to the same group of students in a given year.

B Reading Teacher VA Estimates

In terms of teacher gender, the mean difference (male–female) in teacher VA is -0.009 SD for girls (significant at 1%) and -0.004 SD for boys (significant at 10%). Regarding school type, the mean difference (voucher-public) in teacher VA is -0.006 SD for girls and -0.004 SD for boys, both of which are significant at the 1% level. Figure B.1 displays the density of our reading VA estimates. As can be seen, the differences by teacher gender and school type appear to be negligible.

Figure B.1: Empirical distributions of reading teacher VA estimates by gender and school type



C Alternative normalization of VA estimates

As discussed by Card et al. (2016), while the “sorting” or “between-class” component does not depend on normalization, the within-class component does. Therefore, our results regarding the between-class contribution of teacher VA to the gender gap are invariant to the reference group. On the other hand, the within-class contribution of VA to the gap can be affected by the reference group that is chosen for normalization. Nonetheless, our results and main conclusions are robust to the normalization of VA estimates.

Figure C.1 presents the mean teacher VA estimates separately for male and female teachers in rural and urban schools, along with the corresponding 95% confidence intervals. As seen, male teachers in rural schools have, on average, the lowest VA.

Figure C.1: Teachers' Value-Added, Rurality and Gender

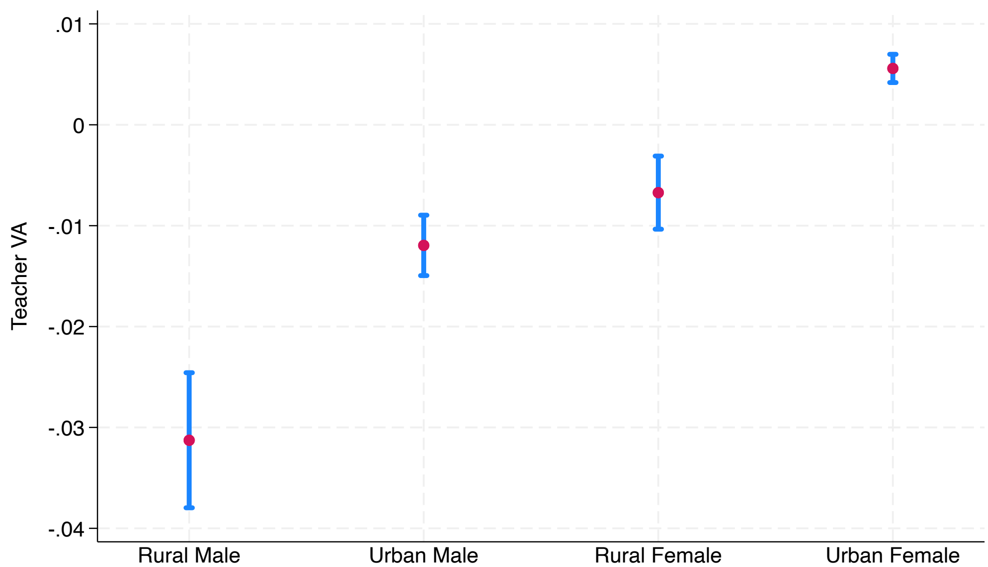


Table C.1 presents the contribution of teacher VA to the gender gap and CCK decomposition when using male teachers in rural schools as the reference group for the normalization of VA measures, with similar findings to those presented in Table 3.

Table C.1: Contribution to Gender Gap and CCK Decomposition using Male Teachers in Rural Schools as Reference Group for the Normalization

					Decompositions of Contribution:			
		Means of TVA		Total	Sorting (between-class)		Gender specific (within-class)	
	Gender	among	among	Contribution	Using Male	Using Female	Using Male	Using Female
	Gap	Boys	Girls	of TVA	Effects	Effects	Distribution	Distribution
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A – Overall Sample:								
Math	0.071	0.038	0.052	-0.014	-0.002	-0.002	-0.012	-0.013
				(-19.7)	(-2.8)	(-2.8)	(-16.9)	(-18.3)
Reading	-0.211	0.017	0.022	-0.005	-0.000	-0.001	-0.004	-0.005
				(2.4)	(0.0)	(0.5)	(1.9)	(2.4)
Panel B – By Teacher Gender:								
Math								
Male	0.095	0.029	0.031	-0.002	0.001	0.001	-0.003	-0.003
				(-2.1)	(1.1)	(1.1)	(-3.2)	(-3.2)
Female	0.059	0.043	0.064	-0.021	-0.004	-0.004	-0.017	-0.018
				(-35.6)	(-6.8)	(-6.8)	(-28.8)	(-30.5)
Reading								
Male	-0.210	0.015	0.015	0.000	-0.002	-0.001	0.002	0.002
				(-0.0)	(1.0)	(0.5)	(-1.0)	(-1.0)
Female	-0.211	0.018	0.023	-0.006	0.000	-0.001	-0.005	-0.006
				(2.8)	(-0.0)	(0.5)	(2.4)	(2.8)
Panel C – By School Dependency:								
Math								
Public	0.064	0.010	0.019	-0.009	-0.001	-0.002	-0.008	-0.008
				(-14.1)	(-1.6)	(-3.1)	(-12.5)	(-12.5)
Voucher	0.093	0.063	0.080	-0.017	-0.001	-0.001	-0.016	-0.016
				(-18.3)	(-1.1)	(-1.1)	(-17.2)	(-17.2)
Reading								
Public	-0.222	0.020	0.026	-0.006	-0.001	-0.001	-0.005	-0.005
				(2.7)	(0.5)	(0.5)	(2.3)	(2.3)
Voucher	-0.193	0.015	0.019	-0.004	0.000	-0.001	-0.004	-0.005
				(2.1)	(-0.0)	(0.5)	(2.1)	(2.6)

Notes: We restrict the analysis to the sample of teachers in mixed classes where teacher VA is estimated for boys and girls, therefore excluding those teachers in single-sex classes. Figures in parentheses represent the percent of the overall gender gap.

D Reading the gender gap: heterogeneity

Table D.1: Contribution to reading gender gap and CCK decomposition, heterogeneous effects

					Decompositions of Contribution:			
		Means of TVA		Total	Sorting (between-class)		Gender specific (within-class)	
	Gender Gap	among Boys	among Girls	Contribution of TVA	Using Male Effects	Using Female Effects	Using Male Distribution	Using Female Distribution
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A.1 – By mother's education, all teachers:								
Primary	-0.196	0.010	0.022	-0.012	0.002	0.001	-0.015	-0.013
				(6.1)	(-1.0)	(-0.5)	(7.7)	(6.6)
Secondary	-0.218	0.013	0.024	-0.011	0.000	-0.000	-0.011	-0.011
				(5.0)	(-0.0)	(0.0)	(5.1)	(5.1)
Tertiary	-0.208	0.015	0.033	-0.018	-0.000	-0.001	-0.017	-0.018
				(8.7)	(0.0)	(0.5)	(8.2)	(8.7)
Panel B.1 – By mother's education, female teacher:								
Primary	-0.196	0.011	0.024	-0.013	0.001	0.000	-0.015	-0.014
				(6.6)	(-0.5)	(-0.0)	(7.7)	(7.1)
Secondary	-0.217	0.013	0.025	-0.012	0.001	-0.000	-0.012	-0.012
				(5.5)	(-0.5)	(0.0)	(5.5)	(5.5)
Tertiary	-0.208	0.014	0.033	-0.019	0.000	-0.001	-0.018	-0.019
				(9.1)	(-0.0)	(0.5)	(8.7)	(9.1)

Notes: We restrict the analysis to the sample of teachers in mixed classes where TVA is estimated for boys and girls; therefore excluding those teachers in single-sex classes. Figures in parenthesis represent the percent of the overall gender gap.