Centre for Economic Performance

Discussion Paper ISSN 2042-2695

No. 1995
April 2024

Teacher value-added and gender gaps in educational outcomes

Andrés Barrios-Fernández
Marc Riudavets-Barcons

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

UKRI Economic and Social Research Council

Abstract
This paper uses rich administrative data from Chile to estimate teacher value added (TVA) on test scores and on an educational attainment index. We allow each teacher to have a different TVA for male and female students and show that differences in TVA explain an important part of the gender gaps we observe in test scores and postsecondary education trajectories. The gaps in gender-specific teaching effectiveness are especially pronounced in mathematics. Indeed, eliminating within-teacher differences in math test score VA would reduce the gender gap in math performance by 67%. We explore what could be behind these gaps in gender-specific TVA and find no significant differences in what makes teachers effective for male and female students. We do find, however, significant associations between teacher characteristics—e.g., gender and performance in the college admission exam—and practices—e.g., paying attention to low-performing students, congratulating students who improve, and having a good relationship with students—with teacher effectiveness. Finally, we also show that math teachers tend to be biased in favor of male students and that teachers with smaller gender biases are more effective for both, male and female students.

Keywords: teacher value-added, education gender gaps, teaching practices
JEL Codes: I21; I24; J24

This paper was produced as part of the Centre's Education & Skills Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We thank the Chilean Ministry of Education, the Education Quality Agency, and the Department of Assessment, Evaluation and Academic Registers (DEMRE) of the University of Chile for granting us access to the administrative data we use in this project. Andrés Barrios-Fernández acknowledges partial support from ANID through FONDECYT grant 11230169 and from the Spencer Foundation through grant 10039719.

Andrés Barrios-Fernández, Universidad de Los Andes, Chile and Centre for Economic Performance at London School of Economics. Marc Riudavets-Barcons, University of Helsinki.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

Teachers are among the most influential actors in children's and teenagers' lives. They not only affect students' academic performance, but also long-term outcomes, including college enrollment and future earnings (Chetty et al., 2014a,b; Jackson, 2018). Despite their relevance, we know little about what makes teachers effective and whether some elements make them better at teaching certain groups of students (Araujo et al., 2016). There is some evidence that male and female students respond differently to certain teachers' characteristics, suggesting that differences in teacher effectiveness across genders could explain part of the gender gaps we observe in educational outcomes (Fryer and Levitt, 2010; Goldin et al., 2006).[1]

This paper provides evidence that teachers are not equally effective at teaching male and female students and that these differences explain an important part of the gender gaps in educational outcomes. To estimate teacher effectiveness, we build on the work of Chetty et al. (2014a) and Jackson (2018). We expand their work by allowing teachers to differentially affect male and female students. Following Jackson (2018), we allow teachers to affect their students in two skills dimensions (i.e., cognitive skills and other skills related to educational attainment).[2] We thus estimate two gender-specific teacher value-added (TVA) indexes for each teacher: one focused on test scores and one focused on educational attainment.

Two contemporaneous papers—Aucejo et al. (2022) and Delgado (2023)—also investigate differences in teacher effectiveness by student types. The former studies differences in effectiveness at teaching students of different genders among language teachers from six urban districts in the United States; the latter studies differences in effectiveness at

[1] Fryer and Levitt (2010) document the emergence of a relevant gender gap in math during primary school. The paper also reports relevant gender gaps both in math and reading test scores among the countries that participate in the Program for International Student Assessment (PISA). Goldin et al. (2006) focuses instead on college attendance. It shows that the gender gap in college attendance and graduation reversed over the past 50 years and that currently, women are more likely than men to attend and complete higher education. The gender gaps we observe in test scores and higher education attendance in Chile—the setting that we study— follow the same pattern documented for the US.

[2] The idea of educators affecting multiple students' dimensions is also used in Mulhern (2023), which studies school counselors' VA and their role in students' higher education trajectories.

teaching students of different races among teachers from public schools in Chicago.[3] As Aucejo et al. (2022), this paper focuses on differences in teacher effectiveness by student gender. However, we implement our analyses in Chile, a setting in which we observe both Spanish and mathematics teachers. Considering that an important part of the public and scientific debate on gender differences in education is centered around math and science, extending previous analyses to math teachers seems important. An additional advantage of the Chilean setting is that it allows us to study teacher effects beyond test scores. We allow teachers to have multidimensional effects—potentially different for female and male students—and study how they impact test scores, but also long-term outcomes such as high school completion and higher education trajectories.

To estimate our TVA measures, we take advantage of rich administrative data that allow us to link students to their eighth-grade math and Spanish teachers, and to follow them throughout high school and in their transition to higher education. Eighth grade is a particularly relevant year for Chilean students as it defines the end of primary education. For most students, this means that they will have to choose a high school and an educational track—i.e., academic or vocational track—at the end of the academic year. This makes eighth-grade teachers particularly relevant for the educational trajectories of their students.

The key challenge for the estimation of a TVA model is to account for any systematic sorting of students to teachers and thus, make sure that differences in outcomes across students capture the causal impact of teachers. To overcome this concern, when estimating our TVA models, we control for prior test scores, as well as for a rich set of socioeconomic characteristics at the student, class, and school levels. Chetty et al. (2014a) shows that controlling for a student's own lagged test scores generates a point estimate of the forecast bias that is small and not statistically significant. In Chile, standardized tests are not applied in every grade. Thus, when estimating eighth-grade teachers' value-added, we can only control for students' test scores from fourth grade. We show

[3] The Online Appendix of Delgado (2023) also presents some results on differences in teacher effectiveness by students' gender. However, the paper focuses on differences in teacher effectiveness by students' race.

that controlling for these test scores and for students' attendance and GPA in each grade up to seventh grade also produces forecast unbiased estimates of teacher effectiveness.[4] We confirm the validity of our estimates by implementing two exercises similar to those presented in Chetty et al. (2014a) that show that our TVA estimates are not affected by student sorting (see Section 6 for further details).

We present three sets of results. Firstly, we show that both TVA indexes—i.e., on test scores and on educational attainment—impact important students' outcomes including contemporaneous test scores, high school graduation, performance in the college admission exam, higher education attendance, and the type of higher education institution attended. Test score TVA is more relevant for students' performance in contemporaneous standardized exams and in the college admission exam, but it still plays an important role in the other outcomes that we study. The opposite is true when focusing on educational attainment TVA. It is less relevant for contemporaneous test scores, but it significantly impacts all the other outcomes that we study. The fact that both TVA measures are relevant for most outcomes confirms that teachers' quality is not uni-dimensional and that there are multiple ways in which teachers influence their students. Our estimates on teacher effects are slightly larger than the ones reported by Jackson (2018), but similar to the ones reported by Chetty et al. (2014a) for the US, by Araujo et al. (2016) for Ecuador, and by Bau and Das (2020) for Pakistan. These last two papers are among the few studies that have estimated TVA in developing countries.

Secondly, we find important differences in teacher effectiveness for male and female students. On average, female students have math teachers with lower test score TVA than male students. We find a difference of 0.12 standard deviations of the test scores distribution (σ_s) in the math test score TVA to which female and male students are exposed. In contrast, female students have teachers with higher Spanish test score TVA ($0.04\sigma_s$) and with higher educational attainment TVA ($0.03\sigma_a$) than male students. Although all these differences are statistically significant, the difference in math test score TVA is by far the

[4] Online Appendix A shows that at least in the Chilean setting there is a strong relationship between GPA and contemporaneous test scores. In addition, in Online Appendix C we show that omitting lagged test scores when controlling for lagged GPA and attendance does not generate major changes to our TVA estimates.

largest one. This result suggests that an important part of the gender gap that we observe in math performance is driven by differences in teacher effectiveness. We then show that the differences in the TVA that female and male students face are mostly explained by within teacher VA differences rather than by student sorting. Following these results, we implement an exercise through which we study how gender gaps in test scores and in higher education attendance would change if we were able to eliminate within-teacher differences in TVA. We find that the gender gap in math test scores would fall by 67%. The gender gap in Spanish test scores and in higher education attendance would also fall but to a lesser extent; they would fall by 15% and 10%, respectively.

Finally, we rely on rich survey data covering the universe of eighth-grade students and math teachers to explore whether teachers' characteristics and practices are associated with their effectiveness. We find a similar association between most teacher practices and their effectiveness at teaching male and female students. This suggests that, at least in terms of the practices that we observe, there are no large differences in what makes a teacher effective for students of different genders. We do not find important differences either in the practices that male and female students report from their teachers, suggesting that teachers are not using dramatically different approaches to teach male and female students. Despite not finding relevant differences in the practices that make a teacher effective for female and male students, we do find some significant associations between teaching practices and teacher value-added. Paying attention to low-performing students, congratulating students who improve, being willing to repeat explanations when asked, and keeping a good relationship with students is positively associated with test score TVA. Some teachers' characteristics also seem to matter. There is a strong association between teachers' test scores in the college admission exam and their value added in test scores. Female teachers are on average more effective at teaching female students. Finally, teachers' gender biases are associated with lower teaching effectiveness for both genders.

In addition to contributing to the literature on teacher value-added, our results add to the research studying the role of teachers on gender gaps in educational outcomes.

An important part of this literature has focused on the effect of a teacher's gender. It has been shown that female students perform better when they have a teacher of their same gender (Dee, 2005, 2007; Carrell et al., 2010; Paredes, 2014; Lim and Meer, 2017). However, Sansone (2017) shows that after controlling for teachers' behaviors, attitudes, and expectations the gender of the teacher does not seem to matter. In the context of higher education, Bettinger and Long (2005) and Porter and Serra (2020) show that female professors influence the major choice of female students. Teachers' gender is not the only characteristic of teachers that has been shown to affect the gender gap. A recent body of evidence shows that teachers' gender stereotypes and biases can also significantly impact students' performance (Carlana, 2019; Lavy and Sand, 2018).

Our work, instead of focusing on specific teachers' characteristics, studies differences in teacher effectiveness across genders. Relying on TVA models we show that differences in teacher effectiveness account for an important part of the gender gaps that exist on educational outcomes. We also study the relationships between teacher effectiveness and a rich vector of teachers' characteristics and practices. Consistently with previous research, we find that female teachers are on average more effective at teaching female students. We also find that teachers with a greater gender bias—measured by discrepancies between students' ranks in standardized test scores and in subject-specific GPA—are less effective at teaching both male and female students. In the setting we study, we show that math teachers are more likely to be biased in favor of male students, which adds to the gap we observe in teacher effectiveness across students' genders.

The rest of the paper is organized into six sections. Section 2 describes the Chilean educational system; section 3 describes the data; section 4 introduces the conceptual framework and empirical approach we use to estimate teacher value-added; section 5 discusses the main results of the paper; section 6 presents two exercises to validate our teacher value-added estimates; and finally, section 7 concludes.

2 Education Institutions in Chile

In Chile, compulsory education lasts 12 years and is organized in two cycles: primary education (grades 1 to 8), and secondary education (grades 9 to 12).[5] After completing their compulsory education, individuals can continue their studies in vocational higher education institutions or universities.

Primary and secondary education is offered by three types of schools: public schools, charter schools, and private schools. Public and charter schools cater to 93% of the students in the country and are currently fully funded by the state through a voucher system.[6] Private schools cater for the additional 7% of the students and are funded through tuition fees. Public, charter, and private schools not only differ in terms of funding, but they are also subject to different regulations and governing bodies, which results in important differences in autonomy (additional details in Barrios-Fernández and Bovini, 2021).

Chile has a nationwide standardized low-stakes exam system (SIMCE). The SIMCE is a multiple-choice exam administered at the end of the school year and marked by external examiners. Students are tested in grades 4, 8, and 10 and although the subjects covered in the exam vary across years, students are consistently tested on math and Spanish.[7]

Higher education is offered by three types of institutions: vocational centers, professional institutes, and universities.

Most universities select their students using a centralized admission system that only considers students' performance in high school and in a national-level university admission exam (PSU). The PSU is taken at the end of the academic year, and since 2006 all students graduating from public and charter schools can register for free. The universities that do not participate in the centralized admission system still have important incentives

[5] Students typically start primary education when they are six years old, and complete secondary education when they are eighteen years old.

[6] Charter schools were able to charge tuition fees until 2016. The resources they received through the voucher system were inversely proportional to the fees they charged.

[7] The testing frequency varies by grade. The SIMCE is applied annually to students in grade 4, and on a regular basis to students in grades 8 and 10.

to consider PSU scores (additional details in Barrios-Fernández, 2021).[8] In contrast, professional institutes and vocational centers do not typically rely on the PSU to select their students.

3 Data

This section describes the sources of the data and the samples we use to estimate teacher value-added and to study its consequences on students' outcomes.

3.1 Data sources

This paper combines administrative data from three public agencies: the Ministry of Education, the Education Quality Agency, and the Department of Evaluation, Assessment and Educational Records (DEMRE) of the University of Chile, the agency responsible for the university admission exam (PSU). We collect data on students, teachers, and schools.

Regarding students, we observe the cohorts starting eighth grade between 2009 and 2014. We follow them throughout primary and secondary education, and in their transition to higher education. In the student registers of the Ministry of Education, we observe the school and class in which they were enrolled, their attendance level, and their GPA from 2002 onwards. We also observe the educational track they choose in high school, and for those who enroll in higher education, the institution and program that they attend. We complement these data with registers from the Education Quality Agency (i.e., the agency in charge of the SIMCE). In the SIMCE registers we observe students' math and Spanish scores in fourth and eighth grades, as well as socioeconomic and demographic characteristics reported by their parents in a survey administered when the students take the SIMCE. Finally, from the DEMRE, we obtain students' performance in the different sections of the university admission exam (PSU), and from the Ministry of Education the higher education institution and program in which students enroll.

[8] For instance, eligibility for most of the financial aid programs for university studies depend on the score students obtain in the PSU.

Regarding teachers, the registers of the Ministry of Education contain information on their gender, age, subject-school-grade-class taught, number of teaching hours, and experience. We complement these data with information on students' perceptions and teachers' practices collected through surveys that both students and teachers answer with the SIMCE.

Finally, from the registers of the Ministry of Education, we gather information on schools and higher education institutions. For schools, we observe their administrative dependence (i.e., public, charter, or private) and municipality, while for higher education institutions, we observe their ownership, location, and years of accreditation.[9] We also observe the duration, tuition fees, and field of study of each higher education program.

3.2 Sample definition

This section describes the sample that we use to estimate teacher value added. To build this sample, we link students taking the SIMCE in eighth grade with their test scores in fourth and eighth grades and with their math and Spanish teachers. We complement these data with the parents' answers to a survey they respond when their children take the SIMCE, and with additional variables from the Ministry of Education and from the DEMRE. By combining these datasets, we create a dataset that includes the universe of grade 8 students taking the SIMCE between 2009 and 2014 in which we observe students' and teachers' links, students' performance in the math and Spanish sections of the SIMCE, their GPA and attendance in eighth grade, whether they enroll and complete high school, whether they register for the PSU, their scores in the PSU, and the higher education program and institution in which they enroll.

Table I presents summary statistics for this sample. The sample is balanced in terms of gender, and the average student is around 14 years old when taking the SIMCE. Around 55% of the students come from households in which monthly income is below CLP 300,000; only 25% of them have a mother who attended postsecondary education;

[9] In Chile, higher education institutions are periodically evaluated by an external authority, the National Accreditation Agency. Based on this evaluation, institutions are awarded with a certificate of quality. This certificate is valid between 2-7 years, depending on how well the institution performs in the evaluation. We define a selective institution as an institution that has been given the certificate for at least 5 years.

and 92% attend a subsidized school. In terms of academic performance, female students have a better GPA and obtain higher scores in the Spanish section of the SIMCE and of the PSU. They are also more likely to complete high school, take the PSU, and enroll in higher education.[10] Male students, on the other hand, perform better on the math section of the SIMCE and of the PSU, and are more likely to enroll in a selective university. We do not observe important differences in the characteristics of the teachers that female and male students have in eighth grade.

We create an additional sample to study the relationship between our measures of teacher effectiveness and teachers' practices. We link each teacher-year in our sample to the teacher value-added measures we estimate, and to the surveys that they and their students answer together with the SIMCE. This allows us to observe additional teachers' characteristics and some of their teaching practices.

4 Teacher Effects

4.1 Conceptual Framework

Teacher value-added models have typically focused on the impact of teachers on students' test scores. It has been shown that having a high value-added teacher in terms of test scores not only improves students' performance on standardized tests, but also increases college attendance and earnings (see for instance Chetty et al., 2014b). However, teachers can also impact their students in other dimensions. Recent work shows that the influence of educators in dimensions not directly related to test scores also plays an important role in shaping educational trajectories (see for instance Jackson, 2018; Mulhern, 2020). We build on these findings and allow teacher effects to be multi-dimensional. Specifically, we allow teachers to differentially impact both, their students' cognitive skills and other skills determining educational attainment.[11] The rest of this section describes

[10] The figures on high school enrollment and completion focus on regular tracks of high school. This means that the actual share of individuals completing high school is larger than the share presented in Table I.

[11] Note that we still allow teachers to influence their students' educational achievement through the formation of cognitive skills. We simply expand the model to allow them to also influence educational

the conceptual framework in which we base our analyses.

Our focus is on eighth-grade teachers. This is an important year for Chilean students as it is the last year of their primary education, after which most of them will have to choose a high school and an educational track.[12] Prior to eighth grade, students acquire a stock of cognitive skills c, and other skills that influence their educational attainment a, described by the vector v_i:

$$v_i = (v_{ci}, v_{ai})^T \tag{1}$$

This initial endowment reflects the cumulative effects of school and other inputs that contribute to its formation, such as family support and extracurricular activities.

In eighth grade, students are assigned teachers for different subjects, including math and Spanish. For simplicity, this conceptual framework focuses on one of them. Teacher j's quality is characterized by the vector w_j. We allow teachers to differentially impact their students' cognitive skills, as well as another set of skills that influence their educational attainment. In addition, since we are interested in studying differences in teacher effectiveness by students' gender, we also allow teachers' effectiveness to differ for male and female students. Thus, we define for each teacher a gender-specific quality vector w_j^G, with $G = \{F, M\}$.

$$w_j^G = (w_{cj}^G, w_{aj}^G)^T \tag{2}$$

Following Jackson (2018), we allow each student to differentially respond to teacher quality, D_i.

achievement through other channels.

[12] There are some schools that offer primary and secondary education in the same establishment, but most subsidized schools specialize in one level of education. In eleventh grade, students can choose between an academic and a vocational track. There are multiple specializations within the vocational track. Not all high schools offer both tracks, and the available specializations also vary across schools.

$$D_i = \begin{pmatrix} D_{ic} & 0 \\ 0 & D_{ia} \end{pmatrix} \tag{3}$$

Thus, the effectiveness of teacher j on student i is given by $w_{ij}^G = D_i w_j^G$, and male and female students' ability at the end of grade eight is defined by the following expression:

$$\alpha_{ij}^G = v_i + w_{ij}^G + \phi_{i-j}^G$$

Where ϕ_{i-j}^G is the impact of the other teachers of student i on his/her ability vector α_{ij}^G. Considering that students' test scores, but also other measures of academic success, including high school completion and enrollment in higher education depend on their ability α_{ij}, we can define the relationship between students' outcomes and their ability vector as follows:

$$Y_{ij} = (\alpha_{ij}^G)^T \beta + \varepsilon_{ij} \equiv (v_i + w_{ij}^G + \phi_{i-j}^G)^T \begin{pmatrix} \beta_c \\ \beta_a \end{pmatrix} + \epsilon_{ij} \tag{4}$$

Where β_c and β_a describe how the outcome Y_{ij} depends on cognitive skills and on other skills related to educational attainment.

Teachers affect students' outcomes through their impact on students' abilities, α_{ij}. From expression (4), the effect of teacher j on students' outcomes is a weighted average of the contribution of the teacher to each ability dimension, $\theta_{ij} = (w_{ij}^{G)^T \beta}$. Therefore, the average effect of teacher j on female and male students is given by:

$$\theta_j^F = \mathbf{E}[w_{ij}^F | F]^T \beta, \text{ and } \theta_j^M = \mathbf{E}[w_{ij}^M | M]^T \beta$$

4.2 Estimation of Teachers' Value Added

To estimate teachers' value-added, we build on Chetty et al. (2014a). However, we allow each teacher to have a differential effect on male and female students. In addition to estimating teachers' effects on test scores, we also estimate their effects on educational attainment. To estimate teacher effects on educational attainment we build an index that combines indicators of high school completion, registration for the university admission exam, enrollment in higher education, enrollment in university, and enrollment in a selective university. All these variables correspond to mid- and long-term outcomes. In Section 5, we show that both dimensions of teacher effectiveness play an important role in shaping students' trajectories in the short- and long-term. Thus, our findings are aligned with Jackson (2018) which also shows that teachers impact their students' outcomes by contributing to the formation of more than one type of skills.[13]

Following the notation introduced in Section 4.1, let i index students, j teachers, and t academic years. In addition, let F_i be a dummy variable indicating whether a student is female. Since we do not observe the vector describing students' initial endowments, we will estimate an empirical version of expression (4):

$$Y_{ijt} = \beta_0 + \beta_1 F_i + \beta_2 X_{it} + \beta_3 F_i \times X_{it} + F_i \theta_{jt}^F + (1 - F_i)\theta_{jt}^M + \varepsilon_{ijt} \tag{5}$$

Where θ_{jt}^F and θ_{jt}^M can be interpreted as the benefits that female and male students respectively receive from teacher j. X_{it} is a vector of students' characteristics that includes past test scores, and socioeconomic and demographic characteristics. In Chile, standardized tests are not applied in every grade. Thus, when estimating the value-added of eighth-

[13] Jackson (2018) estimates teacher contributions to the formation of non-cognitive skills using an index that combines absences, suspensions, GPA, and grade retention. Unfortunately, we do not observe suspensions, and as Online Appendix Figure A1 shows, in Chile course grades are highly correlated with test scores, casting some doubts on whether they are a good measure of non-cognitive skills in our setting. To complement the results in Table III, Online Appendix Table D.I presents results from an exercise in which instead of using our educational attainment TVA, we build an alternative TVA index combining absences and grade retention. The results of this exercise also indicate that teacher effects are multidimensional.

grade teachers, we are only able to control for students' test scores in fourth grade. We show that controlling for these test scores and for students' GPA and attendance in each grade up to seventh grade produces valid estimates of TVA (see Section 6 and Online Appendix C for further details on the validity of our approach).[14] The key identification assumption behind this approach to estimate teacher value-added is that conditional on the set of controls, students' potential outcomes are constant across teachers.

To estimate the teacher value added of teacher j for his/her students on year t, we first obtain $\hat{\theta_{jt}}^{F}$ and $\hat{\theta_{jt}}^{M}$ from expression (5). Then, we predict the value added of teacher j in year t for gender G only using information from the other years in which the teacher j is observed. Thus, the predicted teacher effect is given by the best linear predictor of $\bar{Y}_{jt}^{G}$ based on $\bar{Y}_{j-t}^{G}$:

$$\hat{\theta}_{jt}^{G} \equiv \mathbf{E}[\theta_{jt}^{G} | \theta_{j-t}^{G}] = \psi^{G\prime} \theta_{j-t}^{G} \tag{6}$$

As shown by Chetty et al. (2014a), $\psi^{G} = (\psi_{1}^{G}, ..., \psi_{t-1}^{G}, \psi_{t+1}^{G}, ..., \psi_{t+S}^{G})$ is a shrinkage estimator, where the coefficients of the vector are chosen to minimize the mean-squared error of the forecast of the outcome. By excluding year t from our value-added estimates, we avoid using the same group of students to both estimate teachers' quality and teachers' impact on students' outcomes.

We follow this procedure using test scores and an index of educational attainment as outcomes. Thus, for each teacher and student gender, we estimate two measures of teacher effectiveness. Considering the different nature of the outcomes used to build them, we hope they also capture different dimensions of teachers' quality. We discuss this in more detail in Section 5.

[14] The full set of controls includes mother's educational level, household income, student age, school administrative dependence, class size, the share of female students in the class, whether the school is situated in a rural area, math and Spanish test scores in grade 4, and GPA and attendance each year up to seventh grade. Individual-level controls are also used to build averages at the class and the school level that are included in the specification.

5 Results

This section presents the main results of the paper. It begins by discussing the effect that teachers have on test scores and other educational outcomes for both male and female students. Then, it presents the differences that we find in the quality of the teachers to which male and female students are allocated in grade eight and the implications of these differences for the gender gaps we observe in academic performance. The section concludes by studying associations between teachers' characteristics and practices and their effectiveness at teaching female and male students.

5.1 Teacher Value Added and Students' Outcomes

We study the effect of teachers on different educational outcomes of both female and male students. According to the conceptual framework introduced in Section 4.1, teachers can affect students' outcomes by improving either their cognitive skills or another set of skills that influence their educational attainment. To study this in more detail, we estimate teachers' value added (TVA) in two outcomes: test scores, and an index of educational attainment. Note that both TVA estimates capture a weighted average of teachers' impact on cognitive and educational attainment skills. However, since the outcomes behind each TVA estimate are of different nature, these TVA likely reflect different dimensions of teachers' effectiveness. The correlations presented in Panel (A) of Table II are consistent with this idea; test score TVA and educational attainment TVA are positively correlated, but the correlations are far from one (i.e., 0.19 for math teachers and 0.22 for Spanish teachers).

Thus, to study the effect of teacher effectiveness on different educational outcomes we will rely on the following specification:

$$Y_{ijt}^{G} = \beta_0 + \beta_1 \hat{\theta}_{cjt}^{G} + \beta_2 \hat{\theta}_{cjt}^{G} F_i + \beta_3 \hat{\theta}_{ajt}^{G} + \beta_4 \hat{\theta}_{ajt}^{G} F_i + \beta_5 X_{it} + \varepsilon_{ijt} \tag{7}$$

As shown in expression (7), we include both TVA estimates simultaneously. By including both of them in the same specification, we will be able to capture the effect of different dimensions of teacher effectiveness (i.e., the part that is unique to each TVA index). The superindex G indicates that the same teacher can have different TVA for male and female students. Thus, the TVA measures a student receives in this specification depend on the student's gender. By adding the interaction between the TVA measures and a female indicator, we allow gender-specific teacher effectiveness to differentially affect male and female students. The specification also includes a rich vector of individual, class, and school level controls, and years fixed effects. The causal interpretation of these results relies on the assumption that TVA estimates are orthogonal to unobserved determinants of the outcome conditional on the set of controls that we include.

Table III summarizes the results of this section. Panel A focuses on math teachers, while Panel B on Spanish teachers. We study the effect of teachers' effectiveness in multiple outcomes, including grade 8 test scores, high school completion, performance in the university admission exam, and enrollment in higher education.[15]

TVA estimates are expressed in standard deviations of the student test scores (σ_s) or of the educational attainment index (σ_i) distribution. To make the interpretation of the results on Table III easier, we re-scaled them and they are now expressed in standard deviations of teacher effectiveness in each dimension. β_1, for instance, should be interpreted as the effect that improving test score TVA by one standard deviation of its distribution in the teachers' population (σ_t) has on male students' outcomes (keeping educational attainment TVA constant).Similarly, β_3 should be interpreted as the effect that improving educational attainment TVA by one standard deviation of its distribution in the teachers' population (σ_a) has on male students' outcomes (keeping test score TVA constant).

Our estimates show that in most cases, both dimensions of teacher effectiveness matter. This confirms that our TVA estimates are indeed capturing different dimensions of

[15] We only observe admission exam scores for students who actually take it. As shown in column (3) of Table III teachers also influence the probability of taking the exam. Thus, estimates on admission exam scores should be interpreted with caution.

teacher effectiveness and that in line with Jackson (2018), both dimensions are relevant. According to our results, while test scores TVA is more relevant for students' performance in standardized exams—including the college admission exam—educational attainment TVA is more relevant for outcomes associated with additional years of schooling.

Both teacher effectiveness dimensions impact female and male students to a similar extent. Math teachers' test score TVA is slightly less relevant for female students' mid- and long-term outcomes. Nevertheless, it still significantly impacts their educational trajectories. When focusing on Spanish teachers, there are no significant differences in the impact of test score TVA on students' outcomes by gender. The only exception arises when looking at the probability of attending a selective university, a dimension in which Spanish teachers' test score TVA seems to be more relevant for females.

The effect of math teachers' educational attainment TVA is slightly larger for female students' test scores and for their probability of taking the university admission exam and attending higher education. It is slightly smaller for their probability of enrolling in a STEM degree at university. In contrast, the effect of Spanish teachers' educational attainment TVA is smaller for female students in most outcomes. These differences are statistically significant when looking at their probability of completing high school, taking the college admission exam, attending a selective university, or attending a STEM degree. Note, however, that Spanish teachers' educational attainment TVA still significantly affects all these outcomes for females.

Overall, the results in this section show that both measures of teacher effectiveness significantly impact short- and long-term outcomes for both male and female students. This suggests that differences in teacher effectiveness by student gender might contribute to the gender gaps that we observe in educational trajectories in important ways. We study this in further detail in Section 5.2

In terms of magnitudes, our estimates represent large effects. The differences in TVA that students in the bottom and top third of each TVA distribution face are close to one standard deviations. Thus, the coefficients in Table III can roughly be interpreted as the improvement that students in the bottom 30% of the TVA distributions would experience

if their teachers were replaced by the teachers of the students in the top 30% of the TVA distributions.[16]

These estimates are in line with previous findings. The effect we estimate for test-scores-TVA on test scores is larger than the reported by Jackson (2018), but similar to the reported by Chetty et al. (2014a), Araujo et al. (2016), and Bau and Das (2020). Similar to Jackson (2018), we find that on average, educational attainment TVA does not significantly impact test scores. Nevertheless, we do find that it does impact test scores for female students. Something similar happens when looking at other outcomes. In comparison to Jackson (2018), we find a larger effect of test score TVA on high school completion. However, the effects we find for educational attainment TVA on high school completion are close to the effect that Jackson (2018) finds on this outcome for the behavioral index TVA. Finally, our estimates for the effect of test-scores-TVA on higher education attendance are again larger than those in Jackson (2018), but close to the ones that Chetty et al. (2014a) finds. As in the case of high school completion, our estimates of the probability of attending higher education are similar to the estimates that Jackson (2018) reports for the behavioral index TVA.

5.2 Differences in TVA for Male and Female Students

The results discussed in the previous section indicate that differences in teacher effectiveness can significantly impact short- and long-term educational outcomes of male and female students. In Figure I and Table II, we study differences in the value-added of the teachers to which female and male students are allocated in eighth grade. To build these figures, we use our gender-specific TVA estimates. This means that a female and male student allocated to the same teacher will be exposed to a different TVA.

Panel (a) of Figure I shows that on average female students have math teachers with lower test-score TVA than male students. We find an average difference of 0.118 standard deviations of the test scores distribution (σ_s). According to Table III, improving teacher

[16] The differences that we observe in the TVA that students in the bottom and top third of TVA distribution face in grade eight are: $1.028\sigma_t$ (Δ in math scores TVA), $0.982\sigma_a$ (Δ in math teachers ed. attainment TVA), $0.986\sigma_t$ (Δ in Spanish scores TVA) and $0.982\sigma_a$ (Δ in Spanish teachers ed. attainment TVA.

effectiveness by one standard deviation (σ_t) improves test scores by 0.176 σ_s. Thus, this difference is equivalent to the change we would observe in test scores by raising teacher effectiveness by 0.67 σ_t. As an alternative benchmark, we can use the results in the class size literature. Two iconic studies in this literature are Angrist and Lavy (1999) and Krueger and Whitmore (2001), which find that a one-unit decrease in class size boosts test scores by $0.017 - 0.019\sigma_s$ and 0.048_s, respectively. Thus, the difference we find in math test score TVA for female and male students is equivalent to reducing class size by between 2.5 and 7 students.

The pattern reverts when focusing on Spanish teachers (see panel (b) of Figure I). In this case, the test-score TVA faced by female students is $0.035\sigma_s$ larger than the one faced by male students. Panels (c) and (d) illustrate average differences in educational attainment TVA. In these cases, we find a difference in favor of female students of around $0.025\sigma_i$ both when looking at math and at Spanish teachers. These differences, as shown in Table II are all statistically significant. However, they are considerably smaller than the difference we find in math test score TVA. Considering that an important part of the public and scientific debate on gender differences in education is centered around math and science, this result is important, as it suggests that these gaps are in part driven by differences in how effective math teachers are at teaching female and male students.[17] In Section 5.3 we study in greater detail whether the characteristics and teaching strategies of math teachers can explain some of the differences in their effectiveness at teaching female and male students.

So far we have only looked at averages. Panel (c) of Table II reports as well correlations between female- and male-specific TVA. Although the gender-specific measures of teacher effectiveness are highly correlated, these correlations are far from one. We also report the slope of a linear fit of female-specific TVA on male-specific TVA. The slope is always

[17] To confirm that the gaps that we report in teacher effectiveness by student gender are indeed different from zero we implemented a permutation test. We randomly allocated gender to students and estimate our gender-specific value-added measures 10,000 times. We then computed the average gaps in teacher effectiveness for male and female students in each iteration and compared them with the gaps we obtained when using the actual gender of students. It turns out that the actual gaps are larger than 99.999% of the simulated gaps. Thus, the implied p-values of these exercises are in all cases smaller than 0.00001. See Online Appendix Figure B2 for further details.

smaller than one and we can always reject the null of the slope being equal to one. In Online Appendix Figure B1, we complement the analyses presented in this section by plotting the distribution of test score and educational attainment TVA for female and male students. We test the null hypothesis that the female- and male-specific distributions of TVA are the same and in all cases, we reject the null with p-values lower than 0.0001.

The differences illustrated in Figure I could be driven either by sorting—i.e., male students being allocated to teachers who are better at teaching both male and female students—or by within teacher differences—i.e., teachers being on average better at teaching male than female students. In Figure II, we study these hypotheses by looking at differences in the gender-specific teacher effectiveness to which male and female students are exposed. In panel (a) we focus on math test score TVA. In this case, if the sorting mechanism is the most important, we should find that female students are allocated to math teachers who are worse at teaching both female and male students. However, we find that the math teachers to which female students are allocated are better at teaching female students than the math teachers to which male students are allocated (i.e., the red bar on the left is less negative than the gray bar on the left). It also shows that the math teachers to which female students are allocated are as good at teaching male students as the teachers to which male students are allocated (the red bar on the right is almost identical to the gray bar on the right). This result suggests that sorting is not the main driver of the results in panel (a) of Figure I. Both, the math teachers to which male and female students are allocated are worse at teaching female students, and if anything, the teachers to which female students are exposed are better for them than the teachers to which male students are exposed.

The situation is slightly different when looking at the gender gaps in Spanish test score TVA. As shown in panel (b) of Figure II, male students are allocated to Spanish teachers who are better at teaching them than the teachers to which female students are allocated (i.e., the gray bar to the right is less negative than the red bar to the right). At the same time, male students are allocated to Spanish teachers who are worse at teaching female students than the teachers to which female students are allocated (i.e., the gray bar to

the left is smaller than the red bar to the left). Given the magnitude of the differences between red and gray bars on both sides of the panel, in this case, the gap in the Spanish test score TVA to which students are exposed could be slightly reduced by changing the allocation of students to teachers. However, this would come at a cost, as to achieve this objective, we would need to allocate male and female students to teachers who are worse at teaching them. The decline in the Spanish test score TVA would be a result of making female students lose more than male students. As there are still important within teachers differences in their effectiveness at teaching female and male students, it seems more productive to focus on helping teachers to close that gap.

The case of educational attainment TVA—i.e., panels (c) and (d)—is similar to the case of Spanish test score TVA. The differences are smaller, but it also seems that there is scope to reduce the gap in TVA by reallocating students to teachers. As in the case of Spanish test score TVA, achieving this would result in a decline in the average TVA to which male and female students are exposed. Focusing on closing within teacher gaps in educational attainment TVA seems therefore a better way of tackling these differences.

We conclude this section by studying how the gender differences that we observe in test scores and higher education attendance would change if we were able to eliminate within-teacher gaps on TVA. The red bars in Figure III illustrate the actual differences that we observe between female and male students in math test scores (panel a), Spanish test scores (panel b), and attendance to higher education (panels c and d). The gray bars illustrate how these differences would fall by eliminating differences in TVA. To build the gray bars we start by assigning to each teacher his/her maximum test score and educational attainment TVA. Thus, if the teacher is better at teaching male students, we use his/her male-specific TVA. If the teacher is better at teaching female students, we use his/her female-specific TVA. We then compute the changes in gender-specific TVA between this counterfactual scenario and the one we observe in reality. Combining these changes in TVA with the results in Table III we can predict changes in outcomes for male and female students, and then use them to compute the counterfactual average gap. The results in Figure III show that the gender gap in test scores would fall by 67% in math

and by 15% in Spanish. In the case of higher education attendance, the gender gap would fall by between 8.5% and 13%.

The differences in teacher effectiveness discussed in this section—and especially the ones observed in math test score TVA—seem to play an important role in gender differences in academic performance and educational trajectories. However, eliminating them is not trivial. The next section uses rich survey data on teachers' practices and students' perceptions to explore this issue in more detail.

5.3 TVA and Teachers' Characteristics and Practices

This section takes advantage of rich survey data on teachers' characteristics and practices to study their relationship with our gender-specific measures of teacher effectiveness. We also use this information to explore whether male and female students have different perceptions of what their teachers do in the classroom, something that could shed some light on what is behind the differences we observe in teacher effectiveness. The teacher and student surveys focus on math teachers. Thus, this section will also focus on math teachers. Studying math teachers is important as they are the ones for whom we find the largest gaps in TVA (see Section Section 5.2 for further details).

Differences in teacher effectiveness for female and male students could arise either by teachers using a different approach to teach them or by teaching practices and teachers' characteristics having gender-specific returns.

Firstly, to study the relationship between teacher practices and TVA for male and female students, we estimate a specification in which we regress our gender-specific TVA estimates on teachers' practices and characteristics (X_p):

$$\hat{\theta}^G_{ijt} = \beta_0 + \sum_{p=1}^{P} \beta_p X^G_{pijt} + \varepsilon_{ijt} \tag{8}$$

We run this specification at the student-year level, and in the case of variables recovered from student surveys, we use the answers of female students to explain $\hat{\theta}^F_{ijt}$, and the

answers of male students to explain $\hat{\theta}^{M}_{ijt}$. To distinguish between the two teachers' quality dimensions behind our TVA measures, we use as outcome a residualized version of them. We build these residualized TVA measures by regressing each of them on the other. The goal of this procedure is to keep in each index the part of a teacher value-added that is unique to test scores and to educational attainment, respectively.

The results of this section are summarized in Figure IV. Consistent with previous studies, we do not find a significant association between most teachers' characteristics—i.e., age, years of experience, and hours of contract—and TVA (see for instance Hanushek and Rivkin, 2012; Bau and Das, 2020). We do find, however, that the test score TVA for female students is higher among female teachers than among male teachers. To put this difference in perspective, it represents one-third of the gap that we observe on the test-scores-TVA to which male and female students are exposed to. This finding is in line with previous work that has also found that having a female teacher significantly improves female students' performance in standardized exams (see for instance Lim and Meer, 2017; Gong et al., 2018). Teachers' scores on the university admission exam are also positively correlated with their effectiveness. A difference of one standard deviation in a teacher's university admission exam performance is associated with a difference of around $0.036\sigma_s$ on test score TVA and of around $0.029\sigma_i$ on educational attainment TVA both for male and female students.

The teaching practices that we observe seem to be more relevant for test score TVA. When looking at the information provided by students, we find a positive relationship between test score TVA and paying attention to low-performing students, recognizing students' improvement, repeating explanations when students ask for it, and having a good relationship with the students. These last two characteristics are the only ones that are also positively associated to educational attainment TVA, highlighting the different nature of both measures of teacher quality.

We observe a different set of practices in the teachers' survey. As in the previous case, most of the significant relationships we find arise when focusing on test score TVA. We find that the test-score TVA both for male and female students is positively associated

with conducting interactive classes, asking questions to students, solving exams in the classroom, and frequently using multiple choice exams. This last result is not surprising as standardized exams are also multiple-choice exams. In contrast, we find a negative relationship between test score TVA and making students work in groups too often and relying too much on students' oral expositions.

When turning to educational attainment TVA, we do not find many significant associations with teachers' characteristics and practices. Apart from the association with teachers' university admission exam scores, we only find a positive and significant association with having a good relationship with the students. However, not finding strong correlations between most teacher practices and TVA on the educational attainment index does not necessarily mean that these practices do not influence educational attainment. As shown in Section 5.1, student outcomes are influenced by different dimensions of teacher effectiveness, and test score TVA not only impacts students' performance on standardized exams but also their likelihood to complete high school and attend higher education.

Apart from teachers' gender, the strength of the associations of teacher practices and characteristics with female- and male-specific TVA measures are similar, suggesting that at least in terms of the practices that we observe on the surveys, there are no large differences in what makes a teacher good for students of different genders.

Thus, we next turn to study whether male and female students report different practices among their eighth-grade teachers. Table IV summarizes the results of an exercise in which we regress teachers' characteristics and practices on an indicator of students' gender. Although we find some statistically significant differences, in most cases the estimated coefficients are small. Female students are 3.3 pp less likely to report having a good relationship with their teachers. They are also less likely to report having a teacher who pays attention to low-performing students (0.4 pp), and who congratulates students when they improve (2 pp). In contrast, they are 2.7 pp more likely to have a female teacher. They are also more likely to report having a teacher who repeats explanations when students request it (3.1 pp), who solves problem sets in class (0.6 pp) and who

relies on multiple choice exams (0.5 pp). On average, female students have teachers who performed slightly better than the teachers of male students in the university admission exam (3% of a standard deviation).

Our findings suggest that the difference on teacher effectiveness for female and male students is not explained by differences in these teaching practices. These practices are similarly associated with TVA for male and female students; and although we find some statistically significant differences in how frequently they are used among female and male students, these differences are small.

These results, however, do not imply that teachers' characteristics and practices do not play a role in the gender gaps we observe in educational outcomes. We only observe part of the interactions that occur between students and their teachers, and there is evidence that other dimensions of these interactions, such as gender stereotypes and biases, can generate important differences in students' outcomes (Alan et al., 2018; Carlana, 2019).[18]

Although with the available data, we cannot fully explore the association between teacher-student interactions within the classroom and teaching effectiveness, we conclude this section by conducting an exercise in the spirit of Lavy and Megalokonomou (2019), through which we study the relationship between TVA and the gender-bias of teachers. Since we observe students' test scores and subject-specific GPA, we use this information to build a measure of gender bias at the teacher-year level.

To compute this gender bias index, we first rank students within their classroom according to their test scores and GPA. Next, for each student we compute the difference between these two rankings and normalize the difference by the number of students in the classroom:

$$\Delta R_i = \frac{R_i^{GPA} - R_i^{\text{Test score}}}{\text{Class Size}_i}$$

[18] Psychologists and sociologists have described multiple differences in classroom interactions between teachers and their male and female students. Sadker and Sadker (1985), for instance, argues that teachers pay more attention and give more substantial feedback to male than to female students. Similarly, Dweck et al. (1978) and Rebhorn and Miles (1999) show that teachers are more likely to let female students give up. According to Hyde and Jafeee (1998), math teachers encourage male students to take an independent approach to solving problems, and female students to rely more on predefined rules and computational methods. Finally, Leinhardt et al. (1979) find that at early stages of education, teachers spend different amounts of time teaching reading and math skills to male and female students.

Finally, we build our index of teacher gender bias by computing the difference between the sums of these normalized differences for male and female students:

$$Bias_{jt} = \sum_{i \in M_{jt}} \Delta R_{ijt} - \sum_{i \in F_{jt}} \Delta R_{ijt}$$

Considering that the SIMCE is marked by external reviewers and that the GPA depends on grades decided by the teachers themselves, large positive differences in this index would suggest a teacher grading approach biased in favor of males. Similarly, large negative differences would suggest a bias in favor of females.

We define the 15% of teachers with the smallest bias index in absolute terms as neutral (bias index between -0.03 and 0.03). Teachers with a more negative bias index are defined as pro-female and teachers with a more positive bias index are defined as pro-male. In Figure V we plot the relationship between this teacher bias index and test-scores-TVA for female and male students.

Consistent with Lavy and Megalokonomou (2019), we find that on average neutral teachers have higher test score TVA than pro-female and pro-male teachers. Teacher effectiveness for both male and female students decays with the absolute size of the gender-bias index. However, the speed at which teacher effectiveness decays depends on the sign of the bias. TVA for female students decays faster when we move from neutral teachers towards pro-male teachers; TVA for male students decays faster when we move from neutral teachers towards pro-female teachers. The third row in Table IV indicates that on average math teachers are pro-male and that female students are allocated to teachers who are slightly more pro-male than male students themselves (i.e., we find a positive difference of 0.002 in the gender bias index). These results suggest that teachers' gender bias plays a role in the gender difference we observe in test score TVA. A back-of-the-envelope calculation indicates that by eliminating the gender bias and making all teachers as effective as the neutral teachers, the gap in test score TVA would drop by 7%.

The results discussed in this section suggest that male and female students respond

similarly to different teacher practices and that they are not allocated to substantially different teachers in terms of the characteristics and practices that we observe in our data. Our results, nevertheless, indicate that many teaching practices are associated with high value-added teachers for both genders. Assessing the causal impact of these practices on teacher effectiveness and students' outcomes is a promising avenue for future research.

6 Validation of TVA Estimates

We conclude the paper by presenting two exercises that validate the results discussed in Section 5.1 and confirm that our gender-specific TVA estimates are forecast unbiased. Firstly, we show that our gender-specific TVA estimates are uncorrelated with test scores and educational attainment indexes predicted with variables that do not enter the estimation of TVA. This suggests that sorting on observable characteristics does not seem to be a problem. Secondly, to address concerns related to potential student sorting in unobservable characteristics, we validate our estimates by exploiting quasi-random variation in the pool of teachers that subsequent cohorts of eighth-grade students from the same school face.

6.1 Student Sorting in Observable Characteristics

As noted in Chetty et al. (2014a), an OLS regression of residualized scores on test score TVA should mechanically yield a coefficient of one. Similarly, an OLS regression of the residualized educational attainment index on educational attainment TVA should also yield a coefficient of one. Panels (a) to (d) in Figure VI confirm that this is indeed the case. The relationship between residualized test scores and test score TVA—panels (a) and (b)—and between the residualized educational attainment index and educational attainment TVA—panels (c) and (d)—are first plotted nonparametrically by dividing the relevant TVA estimates into ventiles and then plotting the mean value of the dependent variable independently for mathematics and Spanish teachers. In addition, the panels present a linear fit of the same variables estimated with the underlying microdata at

the student level. The slopes and standard errors of these linear fits are reported at the bottom right corner of each panel.[19]

We find that both test score TVA and educational attainment TVA have a close to one–to–one relationship with residualized test scores and educational attainment index values throughout the distribution. This relationship could be driven by the causal impact of teachers on achievement or by persistent differences in student characteristics across teachers. For instance, our TVA measures may forecast students' test scores in other years simply because some teachers are always assigned to students with more or less educated parents. To study the degree to which the relationship in panels (a) to (d) of Figure VI reflects teachers' causal effects versus bias due to student sorting, we start by estimating forecast bias based on the degree of selection in students' characteristics not included in the VA model. Using data on father education and on the schools that individuals attended between grades four and eight, we predict test scores for each student and then replicate the analyses presented in panels (a) to (d), but using predicted scores or the predicted educational attainment index as dependent variables.

For our baseline TVA specification—which controls for a rich set of prior student-, class-, and school-level scores, attendance, GPA, and demographics and socioeconomic variables—we find that forecast bias from omitting father education and the previous schools attended by students is at most 0.2 percent for test-score TVA and at most 1 percent for the educational attainment TVA. These figures correspond to the top of the 95 percent confidence interval of the estimates reported in panels (e) to (h) of Figure VI. In Online Appendix Figure C.III, we replicate the analyses presented in this section but based on alternative TVA estimates that only control for lagged attendance, GPA, and test scores. This approach allows us to predict test scores and the educational attainment index using a richer vector of variables that includes age, gender, mother's education, father's education, household income, class size, school administrative dependence, and an indicator of whether the school is located in a rural area. As in Figure VI, we find a very

[19] Note that since our TVA measures are gender specific, two students of the opposite gender allocated to the same teacher can have a different TVA associated with them. Thus, for a given year the same teacher might appear at two different points of the x-axis.

flat slope for the relationships between TVA estimates and predicted scores or educational attainment index. The forecast bias from omitting all the sociodemographic variables used to predict scores and the educational attainment index is at most 0.5 percent for test-score TVA and 1.5 percent for educational attainment TVA. Thus, controlling for lagged attendance, GPA, and test scores seems to be enough to obtain valid TVA estimates.

While these results suggest that forecast bias due to sorting on observable predictors of student test scores and educational attainment is minimal, bias due to the omission of other unobservable characteristics could still be relevant. To address this concern, in the next section, we implement a quasi-experimental analysis that confirms the validity of our estimates.

6.2 Teacher Effects and Changes in the Pool of Teachers

This section studies whether student sorting in unobservable characteristics is a threat to the validity of our TVA estimates. We study this by conducting an exercise in which we exploit quasi-random variation in the TVA that subsequent cohorts of eighth-grade students from the same school face induced by changes in teaching staff.

For students enrolled in adjacent cohorts, it is difficult to anticipate and respond to changes in the teachers to which they will be allocated the following year. Since schools still might allocate teachers to specific classes based on some characteristics that we do not necessarily observe, we conduct these analyses at the cohort level. Specifically, we study how changes in average test score TVA across cohorts impact test scores. Since our TVA estimates are gender specific, to implement this exercise we compute average changes in TVA and in test scores independently for female and male students.

Formally, let TVA^{G}_{1st} denote the student-weighted mean of test score TVA (i.e., $\hat{\theta}^{G}_{1jt-\{t,t-1\}}$) across eighth-grade teachers in school s in year t for students of gender G. Similarly, let $\bar{Y_{st}}^{G}$ denote the mean of test scores across eighth-grade students of gender G in school s in year t. We are interested in understanding how changes in TVA^{G}_{1st} impact changes in $\bar{Y_{st}}^{G}$. Thus, we need to build these variables for multiple periods and then compute their differences between consecutive cohorts of eighth-graders. To be sure that changes in test

scores are driven by changes in the pool of teachers and not in the estimated $\hat{\theta}^G_{1jt}$, we build TVA^G_{1st} from $\hat{\theta}^G_{1jt-\{t,t-1\}}$ (i.e., we estimate test score TVA excluding the years that we will be comparing). Therefore, to implement this exercise we need to focus on teachers that we observe teaching eighth grade in at least three periods. We then estimate the following specification:

$$\Delta \bar{Y_{st}}^G = \alpha + \beta_1 \Delta TVA^G_{1st} + \gamma X^G_{st} + \Delta \varepsilon^G_{st} \tag{9}$$

where β_1 is the parameters of interests. It captures the effect of a one-unit improvement in average test score TVA on average test scores. This empirical strategy relies on the assumption that changes in eighth-grade teachers TVA within a school are orthogonal to changes in other determinants of students' outcomes across cohorts (i.e., $Cov(\Delta TVA^G_{st}, \Delta \varepsilon^G_{st}) = 0$).

The results of this exercise are summarized in Figure VII and Table V. Panels (a) and (b) of Figure VII show that changes in test-score TVA closely predict changes in both mathematics and Spanish scores. Indeed, the estimated effect of TVA^G_{st} in both cases is close to one and we cannot reject the null of it being equal to one. In panels (c) and (d) we implement a similar exercise, but we define as a dependent variable the change in the predicted scores described in Figure VI. It is comforting to see that the slope in this case is flat, as it suggests that our identifying assumption holds. Online Appendix Figure C.IV replicates these analyses but using TVA estimates that come from a specification that only controls for lagged attendance, GPA, and test scores. This allows us to predict test scores with a much richer vector of sociodemographic characteristics. The results are remarkably similar to the ones we present in this section, which suggests that controlling for lagged measures of academic performance is enough to generate valid estimates of TVA.

Table V presents a few additional results. Panel (a) focuses on math, while panel (b) on Spanish. The first three columns present different versions of specification (9). The results in the first column come from a version of the specification that only controls for year fixed effects. The specification in the second column adds changes in lagged scores

as controls, and the third column adds on top of that changes in the socioeconomic characteristics used to estimate TVA as controls. In all cases, we find that changes in TVA closely predict changes in scores. The coefficients are always close to one and adding controls does little to the coefficient. In the fourth column, we present the results of the exercise illustrated in panels (c) and (d) of Figure VII. We find that changes in TVA explain very little of changes in scores predicted from father education and from the schools that students attended between grades four and seven. Online Appendix Table C.I provides similar results in which TVA is estimated only controlling for lagged attendance, GPA, and test scores (i.e., the table behind the results presented in Online Appendix Figure C.IV).

7 Conclusions

Teachers play an important role in shaping students' education and life trajectories. Recent evidence has shown that certain teachers' characteristics differentially affect male and female students, suggesting that teacher effects might explain part of the gender gaps that we observe in educational outcomes.

This paper provides evidence that teachers are not equally effective at teaching male and female students and that these differences explain an important part of the gender gaps in educational outcomes. Our results indicate that if we were able to eliminate the differences in teacher effectiveness, the gender gap in test scores would decrease by 67% in math and by 15% in Spanish. Similarly, the gender gap in higher education attendance would fall by around 10%.

Our analyses indicate that the differences that female and male students face in teacher effectiveness are driven by within-teacher differences, rather than by student sorting. Thus, finding ways to close the within-teacher differences in gender-specific TVA is important to tackle gender gaps in academic performance and educational trajectories.

Motivated by these results, we then turn to study whether a rich set of teachers' characteristics and practices explain differences in gender-specific TVA among math teachers.

Most of the characteristics and practices that we study are similarly associated with teacher value-added for male and female students. This pattern suggests that what makes teachers good for female students is similar to what makes them good for male students. In addition, we do not find important differences in the practices that students of different genders report from their teachers, suggesting that at least in the set of practices that we observe, teachers are not using different approaches to teach students of different genders.

These results, however, do not imply that teachers' practices and characteristics do not play a role in the gender gaps that exist in educational outcomes. We only observe part of the interactions that occur between students and their teachers, and there might be relevant aspects of these interactions that we miss in our data.

We do find that female teachers are on average more effective at teaching female students. Consistent with previous research, we also find that gender biases make teachers less effective at teaching both male and female students. In our setting, math teachers are more likely to be biased in favor of male students, which adds to the gender gap we document in teacher value-added.

Finally, we do find significant associations between teachers' characteristics—such as performance in the college admission exam or gender biases—and practices—such as paying attention to low-performing students, congratulating students who improve, being willing to repeat explanations when asked, and keeping a good relationship with students—with TVA for students of both genders. Assessing the causal effect of these characteristics and practices on teacher effectiveness and student outcomes is a promising avenue for future research.

References

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

Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady (2016, 03). Teacher Quality and Learning Outcomes in Kindergarten *. *The Quarterly Journal of Economics 131*(3), 1415–1453.

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

Barrios-Fernández, A. (2021). Neighbors' Effects on University Enrollment. *American Economic Journal: Applied Economics*, forthcoming.

Barrios-Fernández, A. and G. Bovini (2021). It's Time to Learn: School Institutions and Returns to Instruction Time. *Economics of Education Review 80*, 102068.

Bau, N. and J. Das (2020, February). Teacher Value Added in a Low-Income Country. *American Economic Journal: Economic Policy 12*(1), 62–96.

Bettinger, E. P. and B. T. Long (2005, May). Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students. *American Economic Review 95*(2), 152–157.

Borghans, L., B. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). What grades and achievement tests measure. *Proceedings of the national Academy of Sciences 113*(47), 13354–13359.

Carlana, M. (2019, 03). Implicit Stereotypes: Evidence from Teachers' Gender Bias*. *The Quarterly Journal of Economics 134*(3), 1163–1224.

Carrell, S. E., M. E. Page, and J. E. West (2010, 08). Sex and Science: How Professor Gender Perpetuates the Gender Gap*. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a, September). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review 104*(9), 2593–2632.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b, September). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review 104*(9), 2633–79.

Dee, T. S. (2005, May). A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review 95*(2), 158–165.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources XLII*(3), 528–554.

Delgado, W. (2023). Disparate Teacher Effects, Comparative Advantage, and Match Quality. *Annenberg Institute at Brown University*.

Dweck, C. S., W. Davidson, S. Nelson, and B. Enna (1978). Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. *Developmental psychology 14*(3), 268.

Fryer, Roland G., J. and S. D. Levitt (2010, April). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics 2*(2), 210–40.

Goldin, C., L. F. Katz, and I. Kuziemko (2006, December). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives 20*(4), 133–156.

Gong, J., Y. Lu, and H. Song (2018). The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes. *Journal of Labor Economics 36*(3), 743–778.

Hanushek, E. A. and S. G. Rivkin (2012). The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics 4*(1), 131–157.

Hyde, J. S. and S. Jafeee (1998). Perspectives from social and feminist psychology. *Educational Researcher 27*(5), 14–16.

Jackson, C. K. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Krueger, A. B. and D. M. Whitmore (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *The Economic Journal 111*(468), 1–28.

Lavy, V. and R. Megalokonomou (2019). Persistency in teachers' grading bias and effects on longer-term outcomes: University admissions exams and choice of field of study. Technical report, National Bureau of Economic Research.

Lavy, V. and E. Sand (2018). On the Origins of Gender Gaps in Human Capital: Short- and Long-term Consequences of Teachers' Biases. *Journal of Public Economics 167*, 263–279.

Leinhardt, G., A. M. Seewald, and M. Engel (1979). Learning what's taught: Sex differences in instruction. *Journal of Educational psychology 71*(4), 432.

Lim, J. and J. Meer (2017). The Impact of Teacher–Student Gender Matches: Random Assignment Evidence from South Korea. *Journal of Human Resources 52*(4), 979–997.

Mulhern, C. (2020). Beyond Teachers: Estimating Individual Guidance Counselors' Effects on Educational Attainment. *Working paper*.

Mulhern, C. (2023, November). Beyond Teachers: Estimating Individual School Counselors' Effects on Educational Attainment. *American Economic Review 113*(11), 2846–93.

Muñoz, P. and M. Prem (fc). Managers' Productivity and Recruitment in the Public Sector. *American Economic Journal: Economic Policy*.

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.

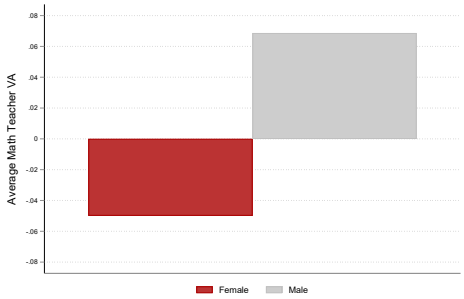
Porter, C. and D. Serra (2020, July). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics 12*(3), 226–54.

Rebhorn, L. S. and D. D. Miles (1999). High-Stakes Testing: Barrier to Gifted Girls in Mathematics and Science? *School Science and Mathematics 99*(6), 313–319.

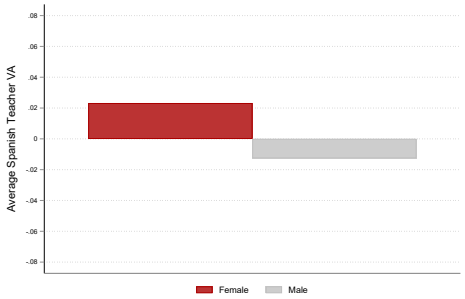
Sadker, M. and D. Sadker (1985). Sexism in the Classroom. *Vocational Education Journal 60*(7), 30–32.

Sansone, D. (2017). Why does Teacher Gender Matter? *Economics of Education Review 61*, 9–18.

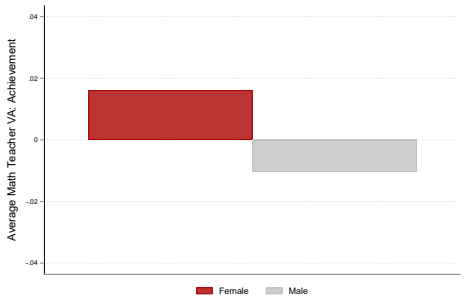
Figure I: Average Gender-Specific-TVA for Female and Male Students



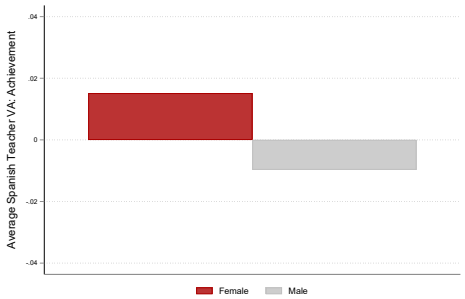
(a) Math Test Scores



(b) Spanish Test Scores



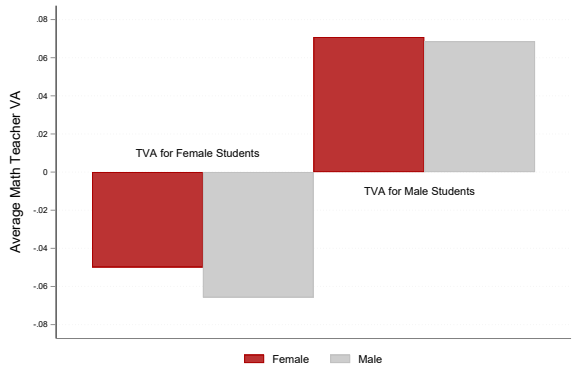
(c) Educational Attainment (M)



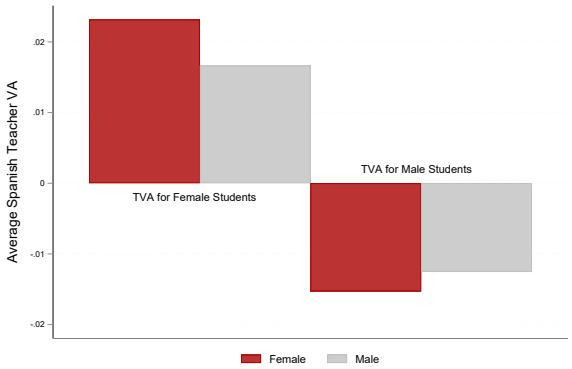
(d) Educational Attainment (S)

Note: This figure illustrates the average TVA to which male and female students are exposed to. Panel (a) presents TVA averages on math test scores, panel (b) TVA averages on Spanish test scores, and panels (c) and (d) TVA averages on the educational attainment index of math and Spanish teachers, respectively.

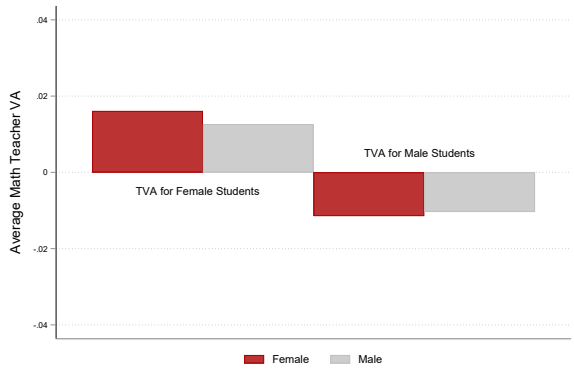
Figure II: Average Female-TVA and Male-TVA by Students' Gender



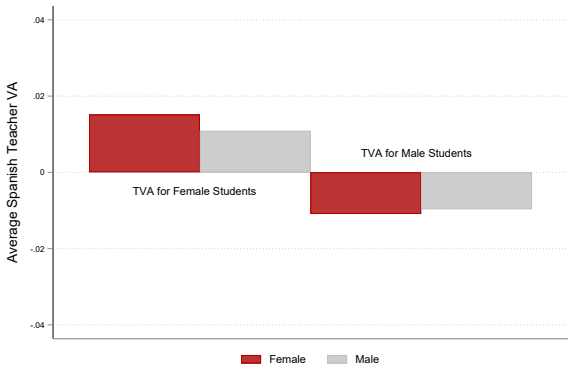
(a) Math Test Scores



(b) Spanish Test Scores



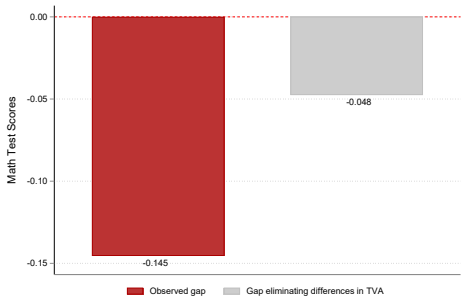
(c) Educational Attainment (M)



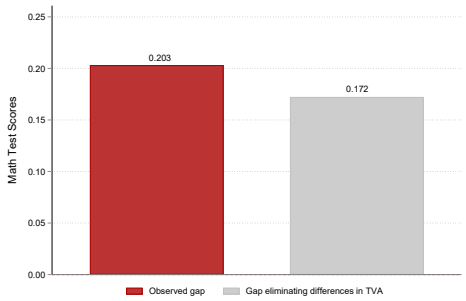
(d) Educational Attainment (S)

Notes: This figure illustrates the average female and male-specific TVA to which both male and female students are exposed to. Panels (a) and (b) present TVA averages in math and Spanish test scores. Similarly, Panels (c) and (d) illustrate TVA averages in the educational attainment index. Panel (c) focuses on math teachers, while Panel (d) on Spanish teachers.

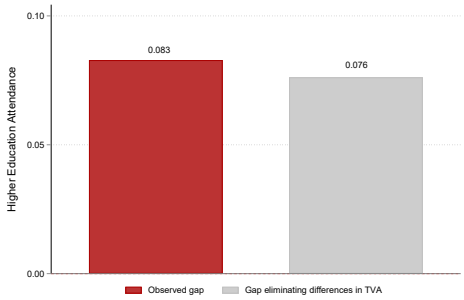
Figure III: Gender Gaps on Educational Outcomes and Gender Differences in TVA



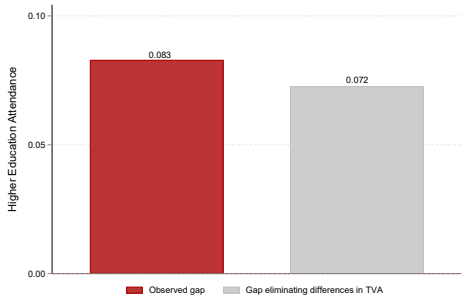
(a) Math Test Scores



(b) Spanish Test Scores



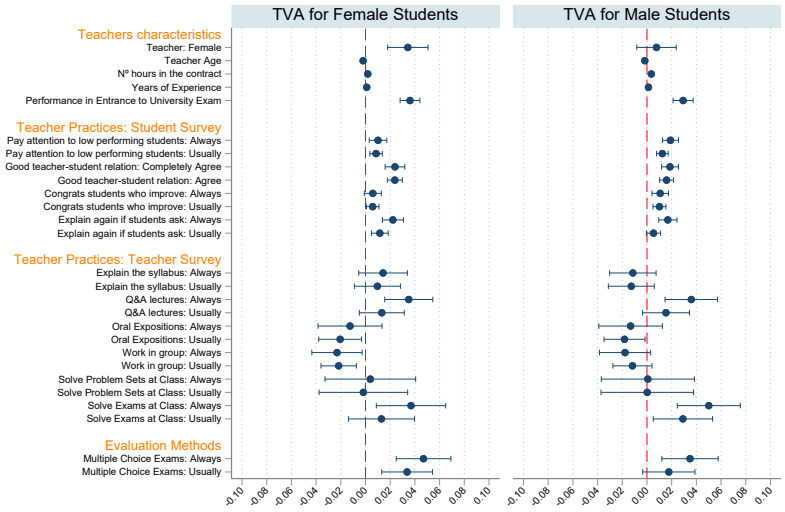
(c) Higher Education Attendance (M)



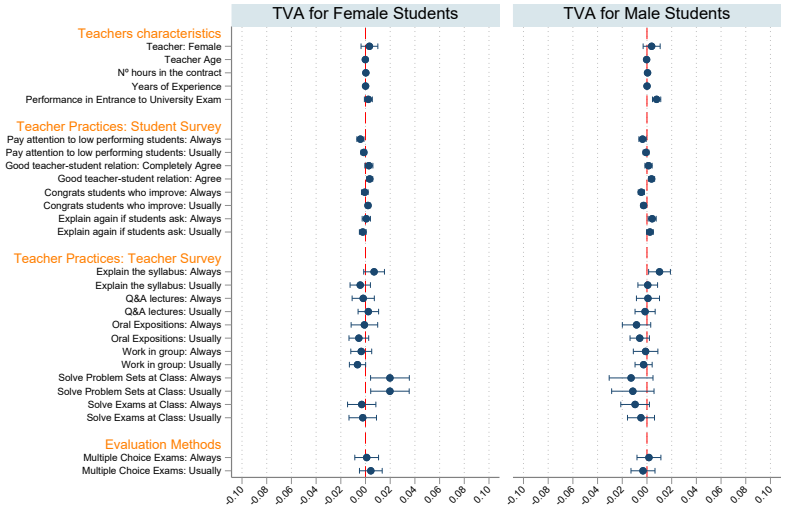
(d) Higher Education Attendance (S)

Note: This figure illustrates the change that the gender gap in test scores and in higher education attendance would experience if we were able to eliminate within teacher gaps on TVA. Red bars represent the gender gap that we actually observe in these outcomes, while grey bars the gap that we would observe in the hypothetical scenario of equal teacher effectiveness. To compute the change in the gap we first assigned to each teacher his/her maximum gender-specific TVA. Combining changes in TVA with the estimates in Table III we then predicted the changes in outcomes that we used to estimate the gap in the counterfactual scenario. See Section 5.2 for further details. Panel (a) focuses on math test scores, panel (b) on Spanish test scores, and panels (c) and (d) on higher education attendance. In all cases, the gap is computed as the difference in the outcome of interest between female and male students.

Figure IV: Gender Specific TVA and Teachers' Characteristics and Practices



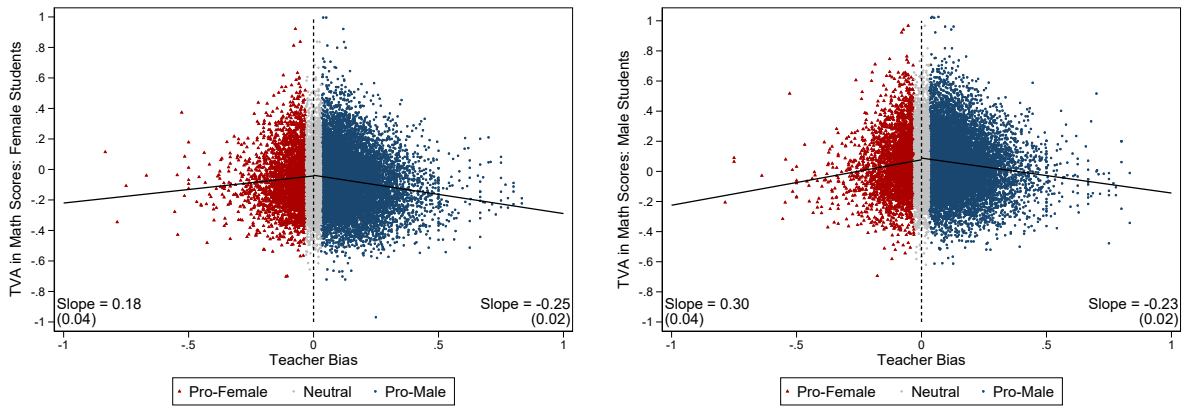
(a) Teacher Value Added in Math Test Scores



(b) Teacher Value Added in Ed. Attainment Index

Note: This figure illustrates the relation between TVA estimates and teachers' characteristics and practices. Coefficients (blue dots) associated with a given characteristic or practice are plotted with their 95% confidence intervals. Standard errors are clustered at the teacher level. The base category of the variables under "Teacher-student interactions"and under "Teacher-practices"is sometimes/never. Panel (a) focuses on TVA on test scores and panel (b) on TVA on the educational attainment index.

Figure V: Teacher Gender-Bias and Value Added in Math Test Scores

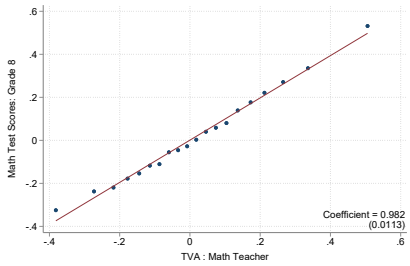


(a) TVA for Female Students

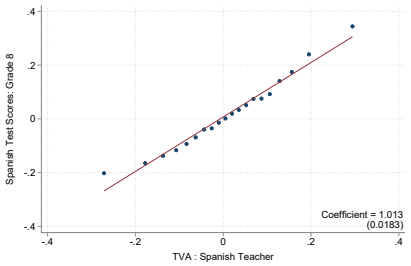
(b) TVA for Male Students

Note: The figures above present scatter plots and linear fits of the math teacher gender-bias index on test-scores-TVA. Panel (a) focuses on test-scores-TVA for female students; while panel (b) on test-scores-TVA for male students. The slope of the linear fits is allowed to change at 0.

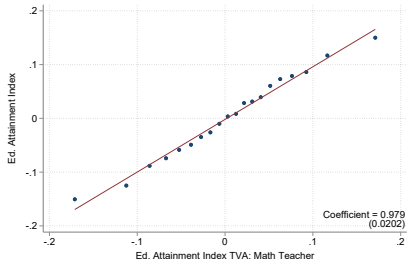
Figure VI: Effects of TVA on Actual and Predicted Scores and Ed. Attainment Index



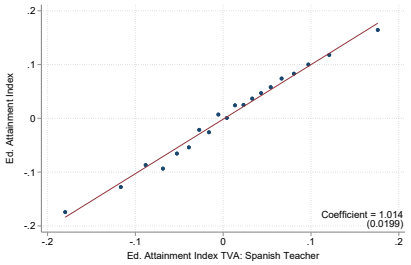
(a) Residualized Math Test Scores



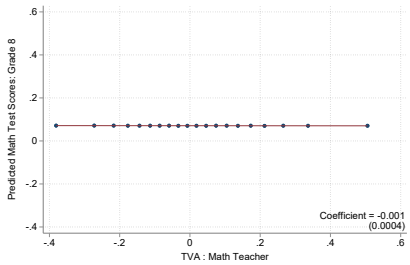
(b) Residualized Spanish Test Scores



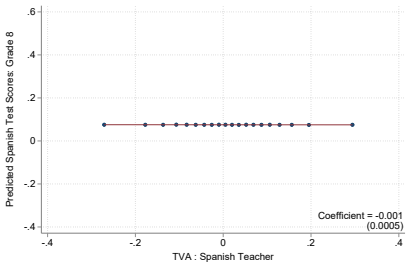
(c) Residualized Ed. Attainment Index (M)



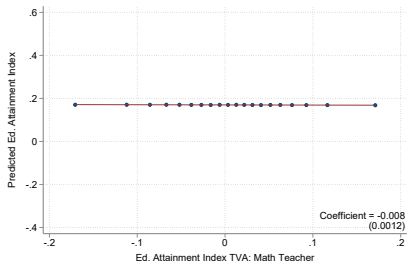
(d) Residualized Ed. Attainment Index (S)



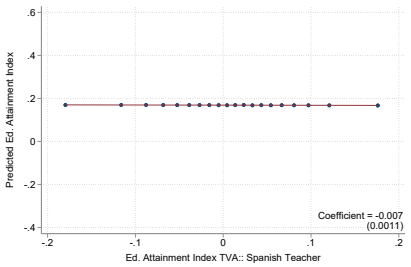
(e) Predicted Math Test Scores



(f) Predicted Spanish Test Scores



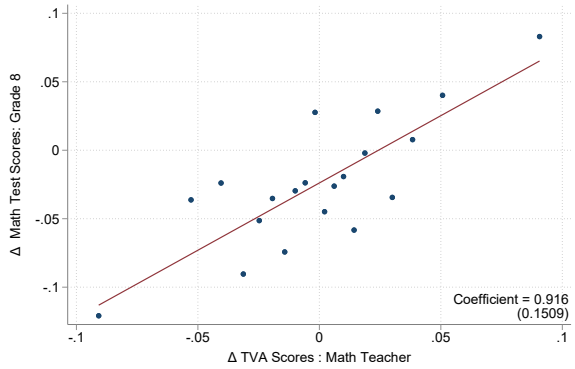
(g) Predicted Ed. Attainment Index (M)



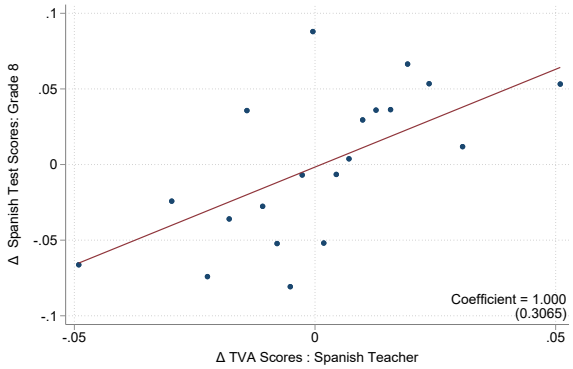
(h) Predicted Ed. Attainment Index (S)

Notes: This figure illustrates the relationship between our teacher value-added estimates and residualized or predicted test scores and educational attainment indexes. Each panel plots these relationships nonparametrically by dividing TVA estimates into ventiles and plotting the mean value of residualized test scores (panels a and b), residualized educational attainment index (panels c and d), predicted test scores (panels e and f), and predicted educational attainment index (panels g and h) independently for mathematics and Spanish teachers. The panels also plot linear fits of these relationships that use the underlying microdata. The slopes and standard errors of these linear fits are reported at the bottom right corner of each figure. We predict test scores and the educational attainment index using father education and indicators of the schools that students attended between grades 4 and 7, all variables not used in the estimation of TVA.

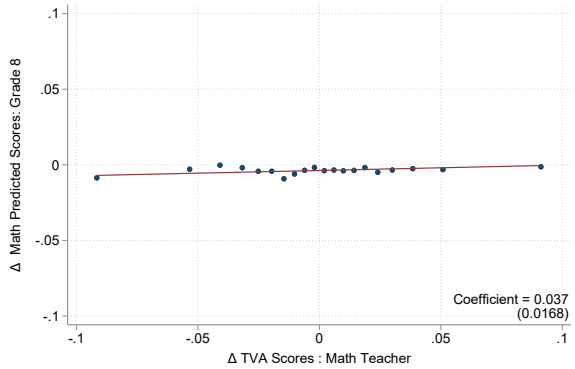
Figure VII: Effects of Changes in Eighth Grade Teaching Staff on Scores across Cohorts



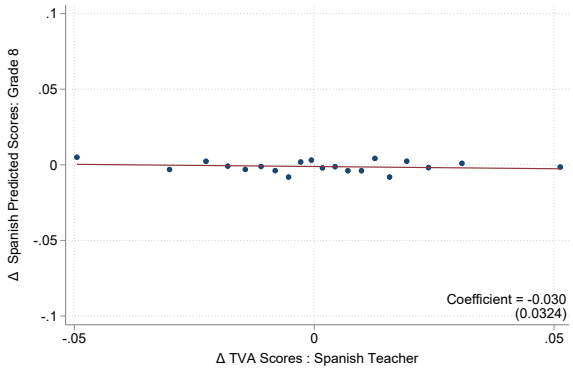
(a) Δ in Actual Math Scores



(b) Δ in Actual Spanish Scores



(c) Δ in Predicted Math Scores



(d) Δ in Predicted Spanish Scores

Notes: This figure illustrates how changes in TVA induced by turnover of eighth-grade Spanish and mathematics teachers affect changes in actual test scores (panels a and b), and in predicted scores (panels c and d) across cohorts. All figures present binned scatter plots of the relationships of interest. The solid line shows the best linear fit estimated on the underlying microdata. The estimated slope of this linear fit, and its standard errors clustered at the school-cohort level are presented at the bottom right corner of each figure. The linear regressions illustrated in the figure correspond to the specifications presented in columns (3) and (4) of Table V. See the notes in Table V for further details.

Table I: Summary Statistics

	All Students (1)	Male Students (2)	Female Students (3)
A. Demographic characteristics			
Gender = Female	0.49	0.00	1.00
Age	13.8	13.9	13.8
B. Socioeconomic characteristics			
High income (> CLP 800k)	0.15	0.15	0.15
Mid income (CLP 300K - 800k)	0.31	0.31	0.30
Low income (< CLP 300K)	0.54	0.54	0.55
Mother education: Less than high school	0.39	0.39	0.40
Mother education: Completed high school	0.34	0.34	0.34
Mother education: Some post secondary	0.27	0.27	0.26
Public School	0.45	0.46	0.44
Voucher School	0.47	0.46	0.48
Private School	0.07	0.07	0.07
Class size	33.19	33.04	33.36
C Academic characteristics			
Attendance in grade 8	0.92	0.92	0.92
GPA in grade 8	5.52	5.44	5.59
Math SIMCE score in grade 8	0.00	0.07	-0.07
Spanish SIMCE score in grade 8	0.00	-0.10	0.10
Enrolls in High School	0.89	0.88	0.90
Graduates from High School	0.74	0.70	0.77
Takes PSU	0.61	0.56	0.65
Math PSU Score	0.00	0.11	-0.10
Spanish PSU Score	0.00	0.01	-0.01
Attends higher education	0.55	0.51	0.59
Attends selective university	0.32	0.33	0.29
Attends STEM program	0.18	0.27	0.08
D. Teachers' characteristics			
Gender = Female	0.58	0.57	0.59
Age	43.77	43.92	43.61
Years of Experience	15.98	16.09	15.88
Hours in the contract	37.20	37.22	37.18
Observations	1,027,154	518,362	508,792

Note: The first column presents summary statistics for the whole sample, while the second and third column focus on male and female students respectively. GPA ranges from 1 to 7. SIMCE and PSU scores are standardized to have mean 0 and standard deviation 1. The figures on high school enrollment and completion focus on regular tracks of high school. This means that the actual share of individuals completing high school is larger than the share presented in the table. STEM programs included are all higher education programs in basic sciences and engineering.

Table II: Teacher Value Added Correlations

	Math Teachers		Spanish Teachers	
	Test Scores (1)	Educational Attainment (2)	Test Scores (3)	Educational Attainment (4)
	A. TVA Correlations			
Cognitive Dimension	1.000		1.000	
Educational Achievement Dimension	0.1863	1.000	0.220	1.000
	B. Average TVA by Gender			
Mean TVA: Male Students	0.068	-0.010	-0.012	-0.009
Mean TVA: Female Students	-0.049	0.016	0.023	0.015
Difference	0.118	-0.026	-0.035	-0.024
	(0.000)	(0.000)	(0.000)	(0.000)
	C. Relationship between TVA measures for Male and Female Students			
Correlation ($\hat{\theta_j}^F$, $\hat{\theta_j}^M$)	0.734	0.355	0.686	0.352
β of linear fit of $\hat{\theta_j}^F$ on $\hat{\theta_j}^M$	0.708	0.358	0.692	0.0.348
	(0.005)	(0.007)	(0.005)	(0.008)
Number of teachers-year	25,808	25,808	26,010	26,010

Note: This table presents different statistics for our TVA estimates. Columns (1) and (2) focus on math teachers, while columns (3) and (4) on Spanish teachers. Panel A presents the correlations between test score and educational attainment TVA estimates. Panel B presents the average test score and educational-attainment TVA to which male and female students are exposed in grade 8. It also presents the differences in the TVA that students of different genders face. Panel C presents correlations between different measures of male- and female-specific TVA. In addition, it presents estimates of the slope of a linear fit of male-specific TVA on female-specific TVA.

Table III: Gender Specific Teacher's Value Added and Educational Outcomes

	Grade 8 test score (1)	Graduates from high school (2)	Takes university admission exam (3)	Score in the admission exam (4)	Attends higher education (5)	Attends selective university (6)	Attends STEM program (7)
				A. Math Teachers			
1. Test Score TVA	0.176	0.008	0.012	0.051	0.011	0.016	0.006
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
2. Test Score TVA × Female	0.001	-0.003	-0.008	-0.006	-0.007	0.002	-0.010
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
3. Educational Attainment TVA	0.003	0.019	0.025	0.014	0.029	0.015	0.019
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
4. Educational Attainment TVA × Female	0.004	0.003	0.004	-0.003	0.007	-0.003	-0.015
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(1)+(2)	0.177	0.005	0.004	0.046	0.004	0.018	-0.005
	(0.004)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
(3)+(4)	0.008	0.020	0.028	0.011	0.036	0.013	0.004
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
				B. Spanish Teachers			
1. Test Score TVA	0.117	0.003	0.003	0.024	0.003	0.004	0.001
	(0.004)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
2. Test Score TVA × Female	0.002	0.001	0.001	0.001	0.001	0.004	-0.002
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
3. Educational Attainment TVA	0.004	0.021	0.031	0.012	0.034	0.019	0.023
	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
4. Educational Attainment TVA × Female	0.005	-0.004	-0.005	-0.002	0.001	-0.005	-0.020
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(1)+(2)	0.119	0.004	0.004	0.024	0.004	0.009	-0.001
	(0.003)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
(3)+(4)	0.008	0.017	0.025	0.010	0.036	0.015	0.003
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Observations	424319	424319	424319	299388	424319	424319	424319

Note: All regressions include year fixed effects. Specifications also control for family income, education of the mother, school administrative dependence, whether the school is in a rural area, age, class size, the share of female students in the class, math and Spanish test scores in grade four, and lagged attendance and GPA between grades four and seven. Individual-level controls are also used to build controls at class-year and school-year levels. Gender-specific teacher effectiveness is based on the leave-out-year TVA estimates discussed in Section 4.2. To make the interpretation of results easier, both TVA indexes are expressed in standard deviations of teacher effectiveness within the teachers' population (i.e., a one-unit increase in TVA corresponds to an improvement of one SD in teacher effectiveness). Panel A focuses on math teachers, while panel B focuses on Spanish teachers. Columns (1) and (4) refer to scores in the subject taught by each teacher. We only observe university admission exam scores for students taking the exam. Thus coefficients in column (4) should be interpreted with caution. Robust standard errors clustered at the teacher level are presented in parentheses.

Table IV: Teacher Practices and Students' Gender

	Gender= Female (1)	Outcome mean (2)
Teacher: Female	0.027	0.581
	(0.004)	
Teacher: Performance in Entrance to University Exam	0.030	0.610
	(0.008)	
Teacher Bias	0.002	0.078
	(0.000)	
Pay attention to low performing students	-0.004	0.514
	(0.002)	
Good teacher-student relation	-0.033	0.399
	(0.002)	
Congrats students who improve	-0.020	0.421
	(0.002)	
Explain again if students ask	0.031	0.663
	(0.002)	
Work in group	-0.000	0.100
	(0.002)	
Explain the syllabus	0.001	0.365
	(0.002)	
Q&A lectures	-0.002	0.367
	(0.003)	
Oral expositions	0.001	0.051
	(0.001)	
Solve problem sets at class	0.006	0.611
	(0.003)	
Solve exams at class	-0.001	0.542
	(0.003)	
Multiple choice exams	0.005	0.423
	(0.003)	
Observations	901873	

Note: All regressions include year fixed effects. Outcome variables take value 1 and 0. In the case of teaching practices, they indicate that a teacher always uses them. Standard errors clustered at teacher level are reported in parentheses.

Table V: Quasi Experimental Estimates of the Forecast Bias

	Δ score	Δ score	Δ score	Δ *scôre*
	(1)	(2)	(3)	(4)
	Panel A - Math Teachers			
Δ in Avg. TVA	0.906	0.915	0.916	0.0368
	(0.166)	(0.153)	(0.151)	(0.0168)
Observations	5362	5362	5362	5229
	Panel B - Spanish Teachers			
Δ in Avg. TVA	0.981	0.989	1.000	-0.0298
	(0.325)	(0.311)	(0.307)	(0.0324)
Observations	4318	4318	4318	4188
Year fixed effects	Yes	Yes	Yes	Yes
Lagged score controls	No	Yes	Yes	Yes
SES controls	No	No	Yes	Yes
Other subject Δ in Avg. TVA	No	No	No	No

Notes: This table presents the results of two exercises. In columns (1) to (3) we study how changes in test score TVA induced by turnover of eighth-grade Spanish and mathematics teachers affect changes in actual test scores. Column (4) presents the result of a similar exercise, but in which the outcome is defined as changes in scores predicted with father education and the schools that individuals attended between grades four and eight. The specification in column (1) only controls for year fixed effects. Column (2) adds changes in lagged test scores as a control, while column (3) adds changes in sociodemographic characteristics of students as controls (i.e., changes in average age and in the share of females, low-income students, and students whose mothers' highest degree of education is high school). Standard errors clustered at the school-cohort level are presented at the bottom right corner of each figure.

Teacher Value-Added and Gender Gaps in Educational Outcomes

Online Appendix

Andrés Barrios-Fernández Marc Riudavets-Barcons

February 28, 2024

Latest Version

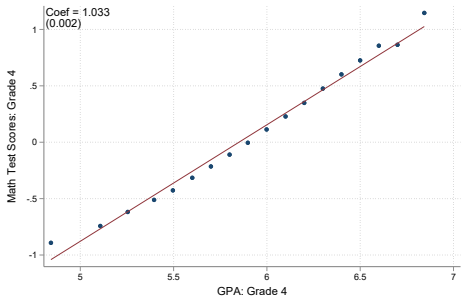
A Relationship between GPA and Contemporaneous Test Scores

This section of the Online Appendix supplements section 4.2 with some evidence on the relationship between GPA and Test-Scores in the Chilean setting. Given that in Chile standardized tests are not applied in every grade, our approach to ensuring that our TVA measures capture teacher effects and not student selection is to include as controls the most relevant measures of students' academic performance that the schools observe between grades four and seven, namely standardized tests in grade four, and grade-specific GPA and attendance between grades four and seven. As discussed in Muñoz and Prem (fc), in the Chilean setting GPA is both highly relevant and informative.

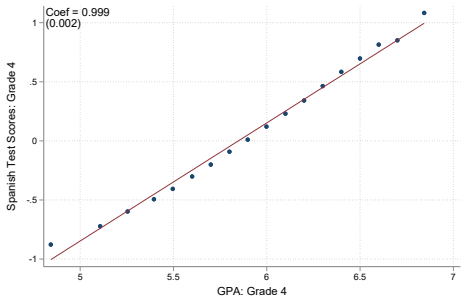
Figure A1, shows that GPA is highly correlated with contemporaneous standardized tests. The figure illustrates the relationship between GPA and test scores in grades four, six, and eight. Although standardized tests are typically applied in grades four and eight, in the period that we study three cohorts of sixth graders were also tested. To give a more comprehensive view of the relationship between GPA and test scores we also used these data to build Figure A1. The slope of the linear regression of fourth-grade test scores on GPA is 1.033 when focusing on mathematics (panel a), and 0.99 when focusing on Spanish (panel b). The same slopes in grade six remain close to one—0.897 when focusing on mathematics (panel c) and 0.899 when focusing on Spanish (panel d). Finally, the same slopes when looking at students in grade eight are 0.827 in mathematics (panel e) and 0.851 in Spanish (panel f). These strong associations between GPA and test scores support the view of using them interchangeably (Borghans et al., 2016).

Figure A1: GPA and Test Scores Correlation

4rth Grade

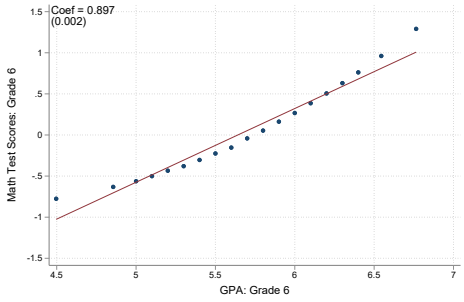


(a) Math

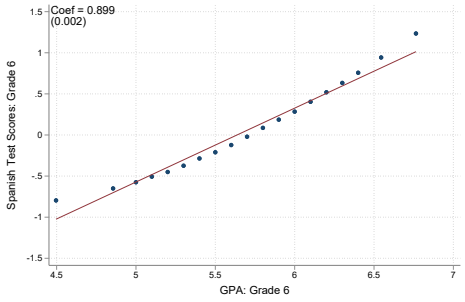


(b) Spanish

6th Grade

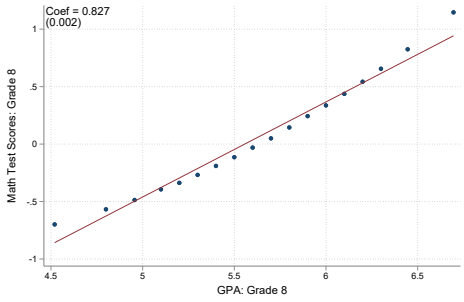


(c) Math

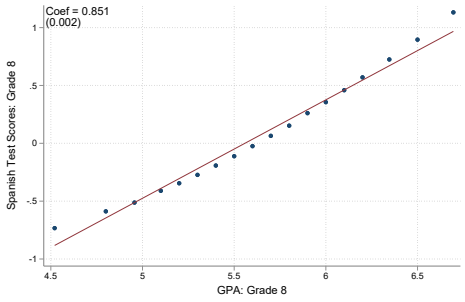


(d) Spanish

8th Grade



(e) Math



(f) Spanish

Note: This figure shows the relation between the SIMCE test scores and the GPA for students in 6th and 8th grade. In panels (a) and (b) we show the relation between average grade GPA and math and Spanish's test scores in 4th grade respectively. In panels (c) and (d) we show the relation between average grade GPA and math and Spanish's test scores in 6th grade respectively. In panels (e) and (f) we show the relation between average grade GPA and math and Spanish's test scores in 6th grade respectively. We report the coefficient and robust standard errors from a linear regression of the test scores on the GPA in the top-left corner on each figure.

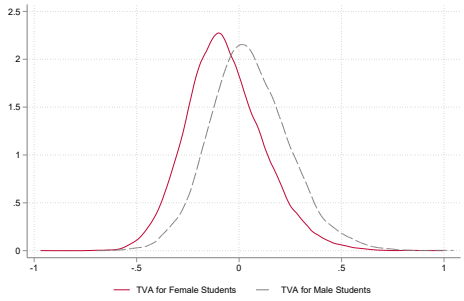
B Distribution of gender-specific TVA Estimates

This section supplements the results in section 5.2 with two additional results on the differences between TVA for male and female students. First, in Figure B1 we plot the the distribution of test score and educational attainment TVA for female and male students. We test the null hypothesis that the female- and male-specific distributions of TVA are the same with the Kolmogorov-Smirnov test-statistic. In all cases, we reject the null with p-values lower than 0.0001.

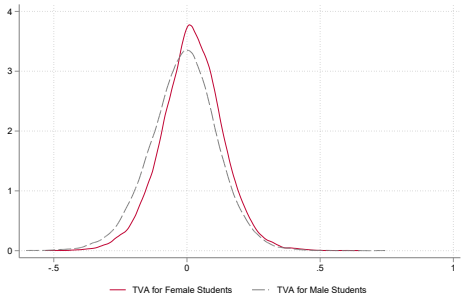
Second, in figure B2 we present a permutation test to confirm that the gaps in teacher effectiveness we report in both Figure I and Table II are indeed different from zero. To do so, we randomly allocate gender to students and estimate our gender-specific value added measures 10,000 times[20]. We then computed the average gaps in teacher effectiveness for male and female students in each iteration and compared them with the gaps we obtained when using the actual gender of students. It turns out that the actual gaps are larger than 99.999% of the simulated gaps. Thus, the implied p-values of these exercises are in all cases smaller than 0.00001

[20] For TVA in math test scores we run this iteration 10,000 times. However, for computer efficiency, for the rest of TVA measures we iterate 1,000 times.

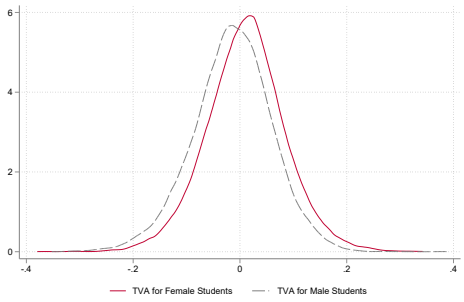
Figure B1: Gender-Specific-TVA Distribution for Female and Male Students



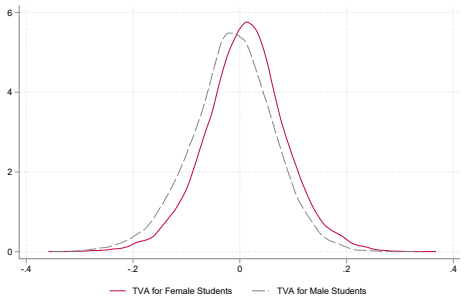
(a) Math Test Scores



(b) Spanish Test Scores



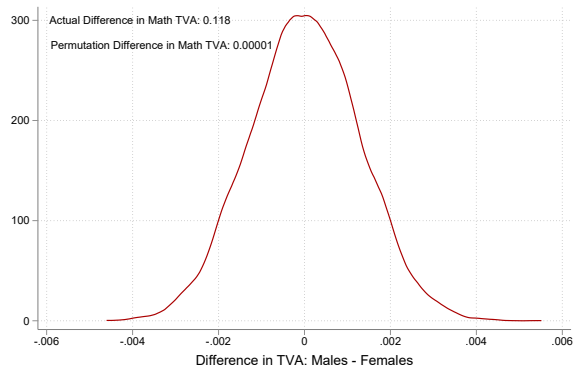
(c) Educational Attainment (M)



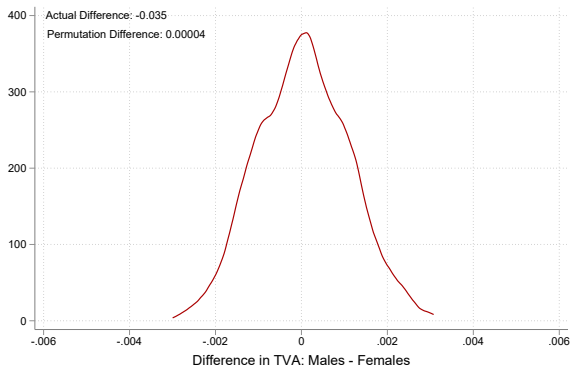
(d) Educational Attainment (S)

Note: This figure illustrates the distributions of the gender-specific TVA to which male and female students are exposed to. Panel (a) presents TVA averages on math test scores, panel (b) TVA averages on Spanish test scores, and panels (c) and (d) TVA averages on the educational attainment index of math and Spanish teachers, respectively.

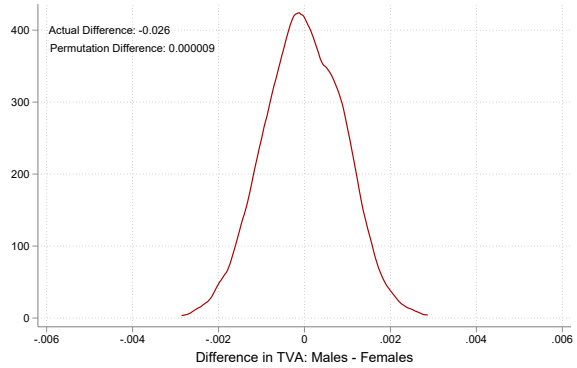
Figure B2: Permutation Exercise



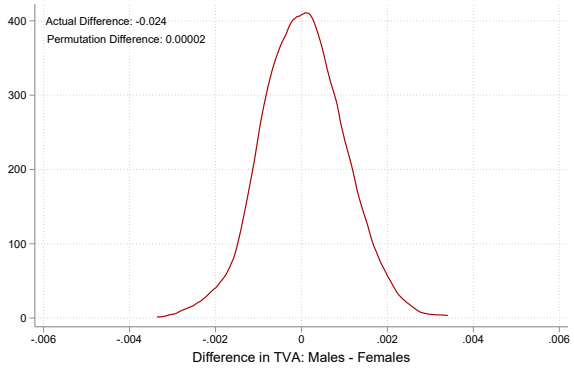
(a) Math Test Scores



(b) Spanish Test Scores



(c) Educational Attainment (M)



(d) Educational Attainment (S

Note: This Figure shows the resulting distributions of the permutation exercise explained in section B. The distributions plot the average gap in teacher effectiveness between male and female students we obtain after conducting all the iterations. Panel (a) focuses on Test-Scores TVA for math teachers, panel (b) focuses on Test-Scores TVA for Spanish teachers, panel (c) focuses on Educational-Attainment TVA for math teachers and panel (d) focuses on Educational-Attainment TVA for Spanish teachers.

C Further Discussion on the Validity of TVA Estimates

This section provides figures and tables that supplement section 6 of the main text, where we discuss the validity of our TVA estimate.

In Figure C.I, we present an additional piece of evidence that confirms that controlling for test scores when GPA and attendance are already included as controls does not significantly change estimates of TVA. Each panel plots the relationship between the measures of teacher effectiveness that we present in the main body of the paper and a parallel measure that we estimated without controlling for past test scores. Hence, we only included lag GPA, lag attendance and the same sociodemographic variables that we use when estimating our main TVA measures. Panel (a) shows that there is a 0.94 rank-rank relation between the two test-score TVA measures for math teachers, while panel (b) shows that the same rank-rank relation for Spanish teachers is 0.91. Panels (c) and (d) do something similar, but focus on educational attainment TVA. The rank-rank relation for math teachers is 0.99, while for Spanish teachers it is 0.87. Finally, in panels (e) and (f) we replicate the previous analyses, but focusing on sixth-grade teachers. For this exercise, we take advantage of the fact that there are three cohorts of sixth graders—i.e., 2013, 2014, and 2015—that were also tested. Although focusing on sixth-grade teachers reduces our sample size, for this sub-sample we can control for lagged test scores from only two years before (i.e., fourth grade). Once more, we find a very high rank-rank relation between estimates coming from specifications that control and omit past test scores (i.e., 0.90 for math teachers and 0.88 for Spanish teachers). TVA estimates that control and not control for lagged test scores are highly correlated, a result that suggests that at least in the Chilean setting, GPA does indeed capture most of the information contained in test scores.

To complement the previous analyses, in Figure C.II we present the average test score value-added of mathematics and Spanish teachers for male and female sixth-graders. Since we observe fourth-grade test scores, by estimating TVA for sixth-grade teachers we

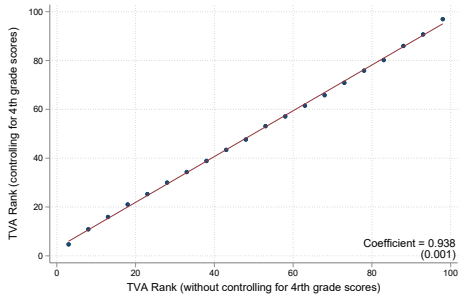
are able to control for two instead of four years lagged test scores. The average TVA and the gaps we find in teacher effectiveness by student gender go in the same direction as those we obtain when focusing on eighth-grade teachers. Specifically, in grade 6 we find a difference in average math teacher Test-Scores TVA gap of 0.047 σ_t in favor of male students and an average difference of 0.025 σ_t in favor of females when focusing on Spanish teachers Test-Scores TVA. This result once more suggests that the findings of the paper are not driven by our inability to control for more recent test scores.

In Figure C.III we test the robustness of our TVA measures by replicating the analysis displayed in Figure VI but constructing a TVA estimate that we obtain when only controlling for lagged attendance, GPA, and test scores. Excluding all available sociodemographic information allows us to predict test scores and the educational attainment index using a richer vector of variables that includes age, gender, mother's education, father's education, household income, class size, school administrative dependence, and an indicator of whether the school is located in a rural area. As in the original exercise, we find a very flat slope for relationships between TVA estimates and predicted scores or educational attainment index. The forecast bias from omitting all the sociodemographic variables that we use to predict scores and the educational attainment index is at most 1 percent for test-score TVA and 2 percent for educational attainment TVA. This suggests that controlling by lagged attendance, GPA, and test scores seems to be enough to obtain valid TVA estimates.

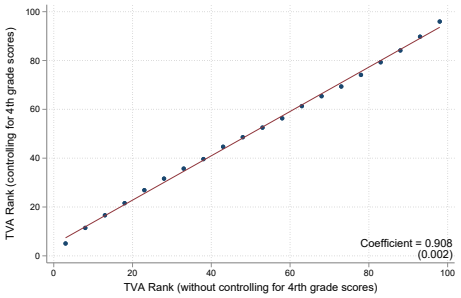
Finally, in this section we conclude showing the results of our quasi-experimental analysis (for details see Section 6.2) when we use the TVA measure we construct only controlling for for lagged attendance, GPA, and test scores. The results of this exercise are summarized in Figure C.IV and Table C.I. Panels (a) and (b) of Figure C.IV show that changes in test-score TVA closely predict changes in both mathematics and Spanish scores. Indeed, the estimated effect of Δ TVA in both cases is close to one and we cannot reject the null of it being equal to one. In panels (c) and (d) we implement a similar exercise, but we define as outcome the change in the predicted scores described in Figure C.III. It is comforting to see that the slope in this case is flat, as it suggests

that our identifying assumption holds. Table C.I presents a few additional results. Panel (a) focuses on math, while panel (b) on Spanish. The first two columns present different versions of the exercise through which we study the impact of changes in TVA over consecutive cohorts of eighth graders on the changes they experience in average test scores. The results in the first column come from a specification that only controls by year fixed effects. In the second column, we add changes in lagged scores as controls. In all cases, we find that changes in TVA closely predict changes in scores. The coefficients are always close to one and adding controls does little to the coefficient. In the fourth column, we present the results of the exercise illustrated in panels (c) and (d) of Figure C.IV. We find that changes in TVA explain very little of changes in scores predicted from a richer vector of variables that includes age, gender, mother's education, father's education, household income, class size, school administrative dependence, and an indicator of whether the school is located in a rural area. Once more, this suggests that our TVA estimates are not capturing changes in students' characteristics.

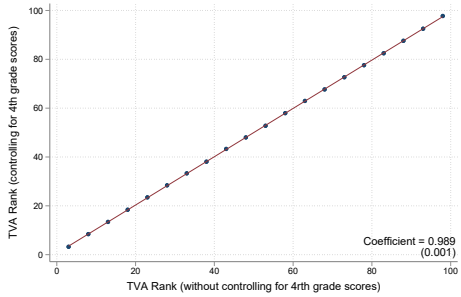
Figure C.I: Rank-Rank Relationship between TVA Estimates with and without Lagged Test Scores



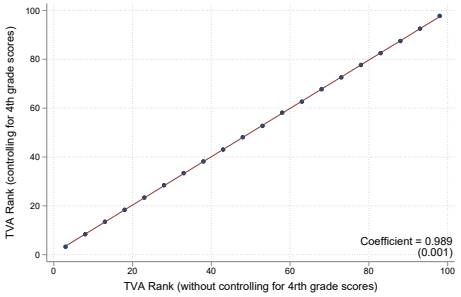
(a) Math Teachers



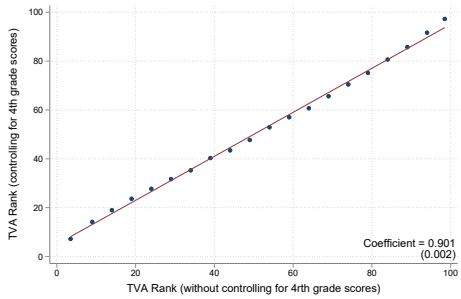
(b) Spanish Teachers



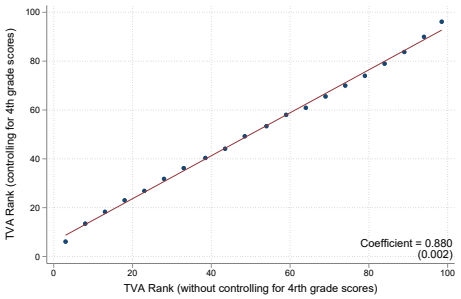
(c) Educational Attainment (M)



(d) Educational Attainment (S)



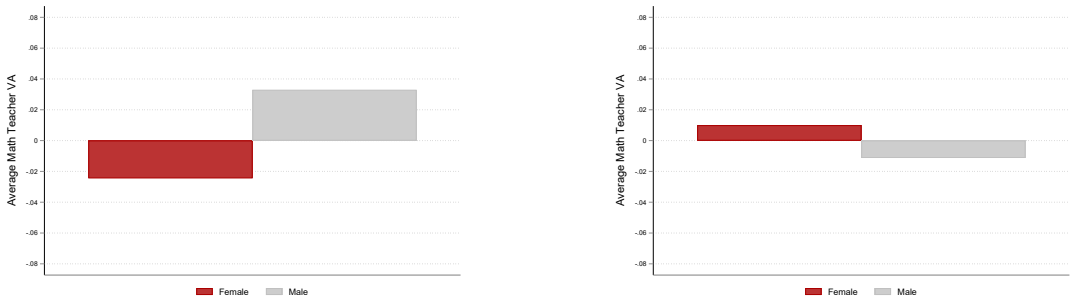
(e) Math teachers (Grade 6)



(f) Spanish Teachers (Grade 6)

Notes: This figure illustrates the rank-rank relationship between TVA measures estimated with and without lagged test scores as controls. All estimates come from specifications in which we control for lagged GPA and attendance. Blue dots correspond to the average rank of teachers according to the value-added measure obtained when controlling for lagged test scores at different levels of the same teachers' rank when not controlling for lagged test scores. The red line corresponds to a linear fit of the same relationship. The slope and standard errors of this linear fit are reported at the bottom right corner of each figure. Panel (a) plots this rank-rank relationship for estimates of math test score value-added, panel (b) for estimates of Spanish test score value-added, panel (c) for estimates of educational attainment value added among math teachers, panel (d) for estimates of educational attainment value added among Spanish teachers, panel (e) plots this rank-rank relationship for estimates of math test score value-added in grade 6, and panel (f) for estimates of Spanish test score value-added in grade 6

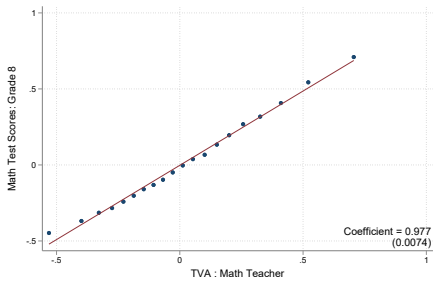
Figure C.II: Gender-Specific Test Score TVA for Sixth Grade Teachers



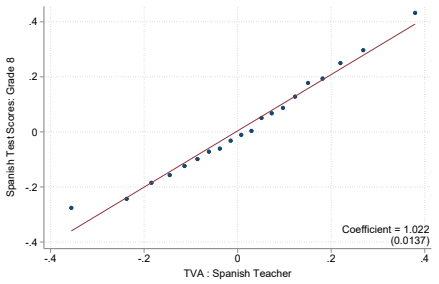
(a) Math (b) Spanish

Notes: This figure plots average test score TVA estimates for sixth-grade teachers. Each teacher is allowed to have a different TVA for female and male students. The red bars illustrate the average TVA for female students, while the gray bars the average TVA for male students. The specification behind these estimates is identical to the main specification presented in the paper. The main difference is that by focusing on sixth-grade teachers, the lagged test scores come from two instead of four years before the date in which we observe teachers. Panel (a) focuses on math teachers, while panel (b) on Spanish teachers.

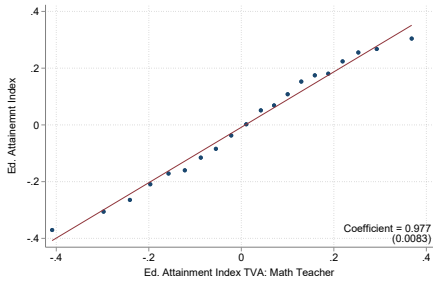
Figure C.III: Effects of TVA on Actual and Predicted Scores and Ed. Attainment Index: TVA build only Controlling by Lagged Attendance, GPA, and Test Scores



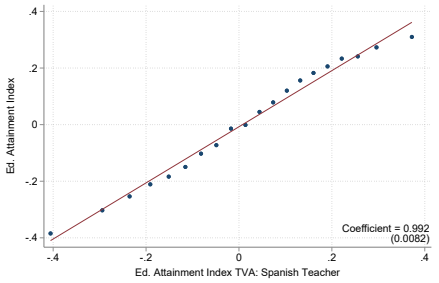
(a) Residualized Math Test Scores



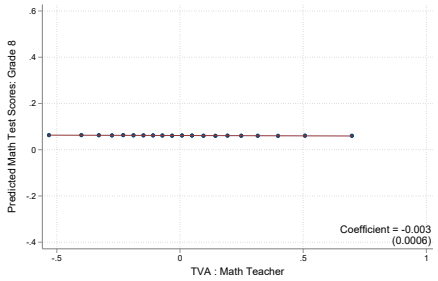
(b) Residualized Spanish Test Scores



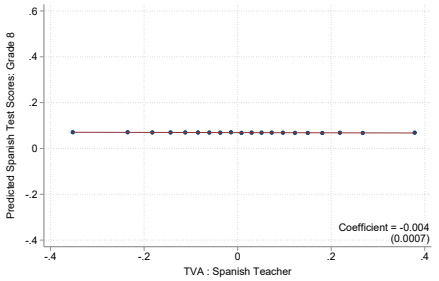
(c) Residualized Ed. Attainment Index (M)



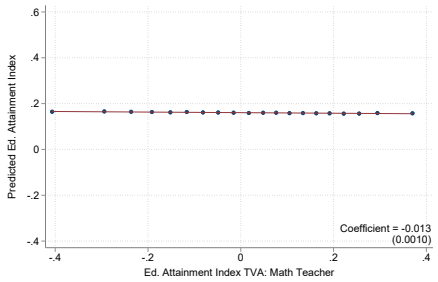
(d) Residualized Ed. Attainment Index (S)



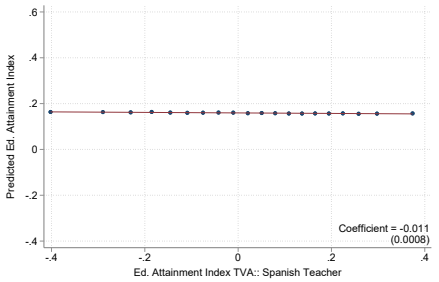
(e) Predicted Math Test Scores



(f) Predicted Spanish Test Scores



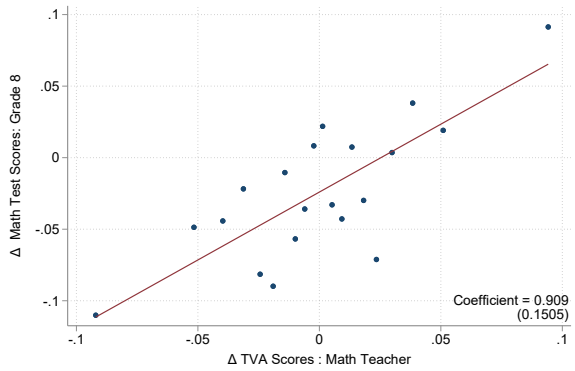
(g) Predicted Ed. Attainment Index (M)



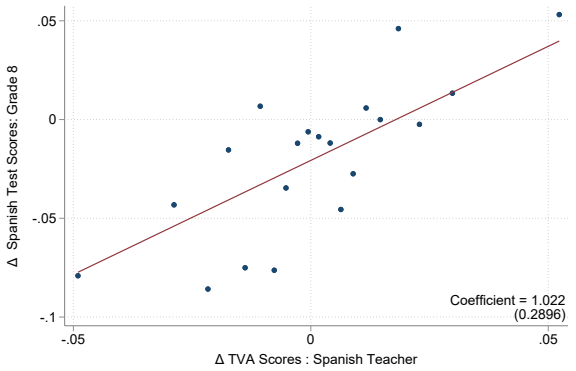
(h) Predicted Ed. Attainment Index (S)

Notes: This figure replicates Figure VI but the TVA measures used here were estimated only controlling for lagged attendance, GPA, and test scores. The predicted scores and educational index used in panels (e) to (h) were built using age, gender, mother's education, father's education, household income, school administrative dependence, class size, and an indicator of whether the school is located in a rural area.

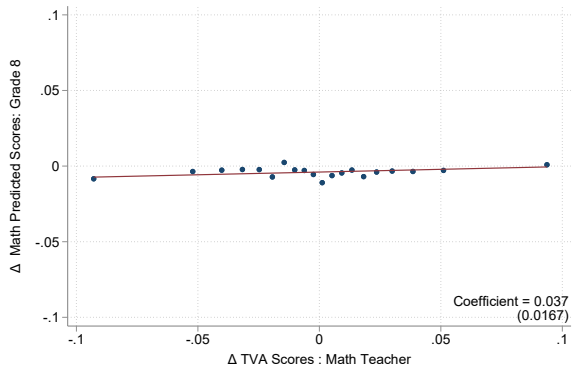
Figure C.IV: Effects of Changes in Eighth Grade Teaching Staff on Scores across Cohorts: TVA Build only Controlling by Lagged GPA and Test Scores



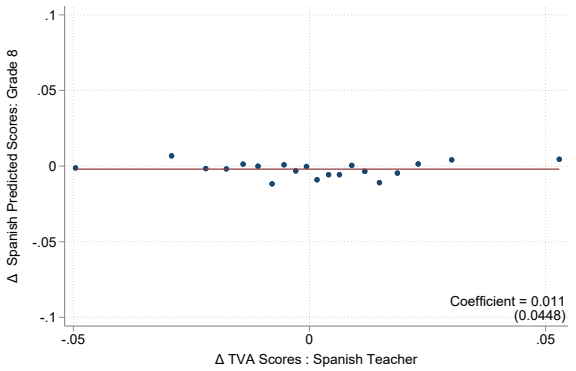
(a) Δ in Actual Math Scores



(b) Δ in Actual Spanish Scores



(c) Δ in Predicted Math Scores



(d) Δ in Predicted Spanish Scores

Notes: This figure illustrates how changes in TVA induced by turnover of eighth-grade Spanish and mathematics teachers affect changes in actual test scores (panels a and b), and in predicted scores (panels c and d) across cohorts. All figures present binned scatter plots of the relationships of interest. The solid line shows the best linear fit estimated on the underlying microdata. The estimated slope of this linear fit, and its standard errors clustered at the school-cohort level are presented at the bottom right corner of each figure. The linear regressions illustrated in the figure correspond to the specifications presented in columns (3) and (4) of Table C.I. See the notes in Table C.I for further details.

Table C.I: Quasi Experimental Estimates of the Forecast Bias: TVA Build only Controlling by Lagged GPA and Test Scores

	Δ score	Δ score	Δ *scôre*
	(1)	(2)	(3)
	Panel A - Math Teachers		
Δ in Avg. TVA	0.947	0.909	0.0371
	(0.164)	(0.150)	(0.0167)
Observations	5362	5362	5362
	Panel B - Spanish Teachers		
Δ in Avg. TVA	1.081	1.022	0.0106
	(0.295)	(0.290)	(0.0448)
Observations	5368	5368	5368
Year fixed effects	Yes	Yes	Yes
Lagged score controls	No	Yes	Yes

Notes: This table presents the results of two exercises. In columns (1) and (2) we study how changes in test score TVA (estimated only controlling for lagged, GPA, and test scores.) induced by turnover of eighth-grade Spanish and mathematics teachers affect changes in actual test scores. Column (4) presents the result of a similar exercise, but in which the outcome is defined as changes in scores predicted with age, gender, mother's education, household income, school administrative dependence, class size, and an indicator of whether the school is located in a rural area . The specification in column (1) only controls for year fixed effects. Column (2) adds changes in lagged test scores as a control. Standard errors clustered at the school-cohort level are presented at the bottom right corner of each figure.

D Additional Results

This section provides results that complement section 5.1. We present an exercise were we substitute our TVA measure in Educational Attainment for a measure of non-cognitive skills TVA, closer to Jackson (2018)[21]. However, contrary to Jackson (2018) in the Chilean setting we do not observe suspensions and, as show in Figure A1 course grades and test scores are highly correlated. Hence, to build our non-cognitive skills index we only use attendance and grade retention. Results of this exercise are presented in Figure D.I and Table D.I.

First, Figure D.I shows the average non-cognitive skills TVA male and female students receive from their maths and Spanish teachers, panels (a) and (b) respectively. In both cases, we find a similar gap in favor of male students of 0.0005 σ_t. However, this gaps are smaller than the ones we find for the educational attainment TVA measure we use in the main text (see Table II). These smaller differences partly reflect the fact that our non-cognitive skills index is build only combining attendance and grade retention in grade 8. At that stage of the education, grade retention is vary rare in Chile and there are no major differences between male and female students. In fact, around 90% of teachers in our sample do not make any of their students repeat grade 8 and overall 99.25% of the students moves to the next grade. In grade 8 attendance is also generally high, above 90% and with no differences between male and female students.

Second, Table D.I presents the results of an exercise similar to Table III where we replace our TVA measure of Educational Attainment for the non-cognitive skills TVA presented in this section. Panel (A) focuses on math teachers and panel (B) focuses on Spanish Teachers. Similar to the results in the main text, these results confirm that teachers impact their students' outcomes not only through contributing to the formation of cognitive skills but also through other paths. Nevertheless, the effects that we find for the non-cognitive skills TVA are smaller than the ones that we find for educational attainment TVA. However, they are still statistically significant for most of the outcomes

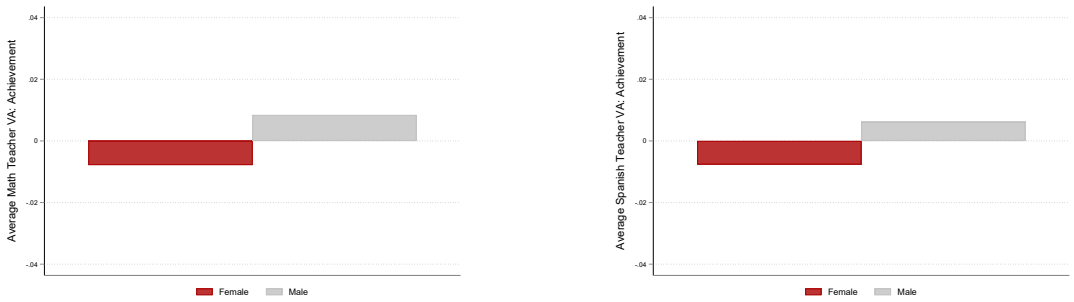
[21] Jackson (2018) builds a non-cognitive TVA measure combining GPA, attendance, suspensions and grade retention

that we study. Finding smaller effects is not entirely surprising given that the non-cognitive skills TVA that we use in this exercise is based only on absences and grade retention. These two variables are unlikely to capture all the non-cognitive skills that impact long-term.

Specifically, the effect of math teachers' non-cognitive skills TVA is slightly larger for female student's probability of taking the university admission exam, having a higher score in it and in the probability of attending higher education and enrolling in a STEM degree. It is smaller for the test scores in grade 8. In contrast, the effect of Spanish teachers' behavioural TVA is smaller for female students in test scores, the score in the admission to university exam and the probability of attending a selective institution in higher educations. It is slightly larger for the probability of attending higher education.

Overall, the results in this section reinforce the results presented in section 5.1 where we concluded that teachers can affect students' short and long run outcomes though multiple channels. The main difference between the non-cognitive skills TVA and the educational attainment TVA that we use in the main body of the paper is that, the former is build using short-term outcomes while the later is based on mid and long-term outcomes that significantly impact students' lives. Given that both indexes seem to point towards the same direction, we believe that the educational attainment TVA allows us to discuss two important points. First, we document a significant, but relatively small gender gap in educational attainment TVA ($0.025\sigma_a$), a result that suggests that focusing on closing gender gaps in test score TVA—especially in math where the gap is $0.012\sigma_s$—should be a higher-order concern (see section 5.2). And second, that TVA measures which capture different teacher's skills are only moderately correlated and hence, there are other relevant dimensions through which teachers impact their students' lives side from the formation of cognitive skills (see section 5.2).

Figure D.I: Gender-Specific Behavioural TVA



(a) Math (b) Spanish

Notes: This figure plots average behavioural TVA estimates for eighth-grade teachers. Each teacher is allowed to have a different TVA for female and male students. The red bars illustrate the average TVA for female students, while the gray bars the average TVA for male students. Panel (a) focuses on math teachers, while panel (b) on Spanish teachers.

Table D.I: Gender Specific Test-Scores and Non-Cognitive Teacher's Value Added and Educational Outcomes

	Grade 8 test score (1)	Graduates from high school (2)	Takes university admission exam (3)	Score in the admission exam (4)	Attends higher education (5)	Attends selective university (6)	Attends STEM program (7)
				A. Math Teachers			
1. Math Test Score TVA	0.176	0.013	0.018	0.054	0.018	0.019	0.011
	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
2. Math Test Score TVA × Female	0.002	-0.005	-0.010	-0.007	-0.009	0.000	-0.015
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
3. Non-Cognitive Skills TVA	0.006	0.002	-0.000	0.004	-0.001	0.004	-0.001
	(0.003)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)
4 Non-Cognitive Skills TVA × Female (D)	-0.003	0.000	0.006	0.002	0.010	-0.000	0.004
	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
(1)+(2)	0.178	0.008	0.008	0.047	0.009	0.019	-0.005
	(0.004)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
(3)+(4)	0.003	0.002	0.006	0.006	0.009	0.004	0.003
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
				B. Spanish Teachers			
1. Spanish Test Score TVA	0.115	0.006	0.008	0.025	0.008	0.006	0.004
	(0.004)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
2. Spanish Test Score TVA × Female (B)	0.004	0.001	0.000	0.001	0.001	0.005	-0.006
	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
3. Non-Cognitive Skills TVA	0.015	0.001	0.000	0.009	-0.000	0.006	0.000
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
4 Non-Cognitive Skills TVA × Female (D)	-0.003	-0.001	0.003	-0.000	0.006	-0.004	0.004
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(1)+(2)	0.119	0.007	0.008	0.026	0.010	0.011	-0.002
	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
(3)+(4)	0.012	0.001	0.004	0.003	0.008	0.002	0.004
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Observations	424319	424319	424319	299383	424319	424319	424319

Note: All regressions include year fixed effects. Specifications also control for family income, education of the mother, school administrative dependence, whether the school is in a rural area, age, class size, the share of female students in the class, math and Spanish test scores in grade four, and lagged attendance and GPA between grades four and seven. Individual-level controls are also used to build controls at class-year and school-year levels. Gender-specific teacher effectiveness is based on the leave-out-year TVA estimates discussed in Section 4.2. To make the interpretation of results easier, both TVA indexes are expressed in standard deviations of teacher effectiveness within the teachers' population (i.e., a one-unit increase in TVA corresponds to an improvement of one SD in teacher effectiveness). Panel A focuses on math teachers, while panel B focuses on Spanish teachers. Columns (1) and (4) refer to scores in the subject taught by each teacher. We only observe university admission exam scores for students taking the exam. Thus coefficients in column (4) should be interpreted with caution. Robust standard errors clustered at the teacher level are presented in parentheses.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1994	Sara Calligaris Chiara Criscuolo Luca Marcolin	Mark-ups in the digital era
1993	Nikhil Datta Stephen Machin	Government contracting and living wages > minimum wages
1992	Antonella Nocco Gianmarco I.P. Ottaviano Matteo Salto Atushi Tadokoro	Leaving the global playing field through optimal non-discriminatory corporate taxes and subsidies
1991	Maria Cotofan Karlygash Kuralbayeva Konstantinos Matakos	Global warming cools voters down: How climate concerns affect policy preferences
1990	Michael Ball Paul Cheshire Christian A.L. Hilber Xiaolun Yu	Why delay? Understanding the construction lag, aka the build out rate
1989	Gianmarco I.P. Ottaviano Davide Suverato	Fantastic beasts and where to find them
1988	Kohei Takeda Atushi Yamagishi	The economic dynamics of city structure: Evidence from Hiroshima's recovery
1987	Gustave Kenedi	Beyond the enrolment gap: Financial barriers and high-achieving, low-income students' persistence in higher education
1986	Stephen J. Redding Daniel M. Sturm	Neighborhood effects: Evidence from wartime destruction in London
1985	Tom Kirchmaier Ekaterina Oparina	Under pressure: Victim withdrawal and police officer workload

1984	Melanie Arntz Sebastian Findeisen Stephan Maurer Oliver Schlenker	Are we yet sick of new technologies? The unequal health effects of digitalization
1983	Javad Shamsi	Immigration and political realignment
1982	Ekaterina Oparina Christian Krekel Sorawoot Srisuma	Talking therapy: Impacts of a nationwide mental health service in England
1981	Xiao Chen Hanwei Huang Jiandong Ju Ruoyan Sun Jialiang Zhang	Endogenous mobility in pandemics: Theory and evidence from the United States
1980	Alan Manning Graham Mazeine	Should I stay or should I go? Return migration from the United States
1979	Rosa Sanchis-Guarner José Montalbán Felix Weinhardt	Home broadband and human capital formation
1978	Luca Macedoni John Morrow Vladimir Tyazhelnikov	Firms in product space: Adoption, growth and competition
1977	Gonzalo Nunez-Chaim Henry G. Overman Capucine Riom	Does subsidising business advice improve firm performance? Evidence from a large RCT
1976	Julian Alves Bruno Serra Jason Greenberg Yaxin Guo Ravija Harjai John Van Reenen	Labour market power: New evidence on Non-Compete Agreements and the effects of M&A in the UK

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE