Teacher Quality and Learning Inequality*

Macarena Kutscher† Catalina Morales‡ Cristina Riquelme§
Sergio Urzúa¶

November 3, 2023

"The art of teaching is the art of assisting discovery",
Mark Van Doren.

Abstract

Schools are expected to equip students with the skills to climb the socioeconomic ladder. This paper examines the contributions of teachers to this process. We explore the determinants of Chile's college admission test for the universe of test takers between 2013 and 2021. The analysis exploits unique and rich matched teacher-student data gathered from multiple administrative information sources, allowing us to account for student, school, and teacher characteristics. We implement different cognitive achievement production function decompositions. Our findings show that teachers' performance on the college admission test and whether they hold a formal educational degree predict their students' success. When switching the focus to the gaps in college admission test scores by school types (public vs. voucher), we document that controlling for students' pre-high-school performance and predetermined school characteristics significantly reduces the average advantage of voucher schools. However, productivity differences emerge as drivers of disparities across the distribution of test scores. Teacher-student interactions play an important role.

*This paper has been prepared for the conference "Aniversario N°50 Revista Estudios de Economía" organized by the School of Economics and Business of the University of Chile. We thank Valentina Paredes for her insightful comments and suggestions.

†Inter-American Development Bank. e-mail: macarenak@iadb.org.

‡University of Maryland. e-mail: cmorale5@umd.edu.

§University of Maryland. e-mail: riquelme@umd.edu.

¶University of Maryland and NBER. e-mail: surzua@umd.edu.

This research used the SIMCE databases of the Chilean Ministry of Education as a source of information. The authors thank the Ministry of Education for access to the information. All the results of the study are the responsibility of the authors and do not compromise said Institution in any way. We thank the Department of Evaluation, Measurement and Educational Registration (DEMRE), of the University of Chile, for providing the databases of the University Higher Education Admission System for the development of this research. This paper was presented in seminars at the Universidad de Chile, The State University of New York at Buffalo, and Duke University. We thank the participants for many helpful comments.

1 Introduction

Education is universally recognized as a crucial determinant of personal development and societal progress (Heckman 2000) and teachers are singled out as a critical input in this process (Rockoff 2004; Rivkin, Hanushek, and Kain 2005; Aaronson, Barrow, and Sander 2007; Chetty, Friedman, and Rockoff 2014b, 2014a; Jackson 2018; Gilraine and Pope 2021; Petek and Pope 2023). Furthermore, research has consistently shown that the learning environment and resources provided by schools have a profound impact on student outcomes (Jackson, Johnson, and Persico 2016). Nonetheless, understanding the complex dynamics between schools, teachers, and student achievement remains an important question with implications for policy development and educational reform.

In this paper, we study the separate factors that influence student achievement, with an emphasis on the impact of teachers and schools. We focus on a critical academic outcome: the performance of students in college admission tests.[1] For this, we exploit a unique, matched school-teacher-student dataset gathered from multiple sources of administrative information within the Chilean educational system. Importantly, our dataset not only comprises extensive student, teacher, and school records but also introduces a novel dimension by incorporating details about teachers' own performance in the high-stakes college admission assessments (at age 17-18), how high they ranked education as their major choice when applying to college, their high school GPA, and detailed information on their education degree.

We investigate whether there are differences among schools' capacity to enhance student academic performance and to what extent these can be attributed to differences in teacher quality. To do this, we employ a multi-step approach. First, we estimate the production function for cognitive achievement, using a value-added specification. Subsequently, we delve deeper to examine whether teacher quality can account for the performance disparities observed across different types of schools. To achieve this, we employ both classical and RIF Oaxaca-Blinder decompositions. Finally, based on the methodology outlined in Firpo, Fortin, and Lemieux 2018, we implement an empirical strategy that considers different distribution moments (unconditional and conditional means and quantiles). we decompose the achievement gap into a *composition* effect (differences in the distribution of observed characteristics) and a *structure* effect (differences in the productivity of observed characteristics). Importantly, we apply this analysis not only to mean performance but to the entire distribution of the performance gap.

The analysis holds significant relevance in the Chilean context, given the pronounced levels

1. For a more structural analysis of the teacher-student relationship and the process of accessing higher education in Chile see Montaño et al. 2023.

of segregation within its educational system and the substantial disparities in student outcomes linked to the type of school they attend. While previous studies have shed light on the influence of socioeconomic factors, a comprehensive examination of the role of teachers is still lacking (Mizala and Romaguera 2000; Contreras 2002; Bravo, Mukhopadhyay, and Todd 2010; Iturra and Gallardo 2022). Therefore, this article aims to go deeper into the impact of teachers. Given the many-to-many nature of the student-teacher relationship and since our analysis focuses on the performance in college admission tests, we focus on high school students. Hence, our student sample comprises 537,119 test takers between 2013 and 2021. An important sample-bound limitation is that we only have access to information on college admission assessments from 2006 onward. Consequently, our analysis is confined to examining the impact of young teachers.

This research yields several findings. First, the value-added model shows that a teacher's own performance in the college admission tests (at age 17-18) has a significant impact on student PSU performance. The same is true for teacher certification, as having a higher proportion of teachers with a formal education degree also yields higher student performance. However, even after controlling for student SES characteristics and a comprehensive set of teacher characteristics, the type of school attended in high school still plays an important role in explaining PSU performance.

Then, we delve into estimating the decomposition of the school-type gap in the average PSU performance by the contribution of different components, such as socioeconomic characteristics, teacher quality, past performance, and propensity to take the PSU test. The Oaxaca-Blinder decomposition implemented using value-added specifications at the mean of the college admissions test score distribution shows that we can explain 15 out of the 23 points gap in math and 12 out of the 28 points gap in Spanish solely by differences in the distribution of the characteristics between the two groups. When we look at the RIF Oaxaca-Blinder decomposition, it shows that the influence of teachers comes prominently in the form of a price effect, indicating significant differences in the productivity of teachers by school type, especially in math. This suggests that teachers with similar characteristics exhibit greater effectiveness in public or voucher schools, contributing to the performance disparity.

When examining the role of teachers throughout the entire distribution of test scores, we identify the heightened significance of the structure effect in the high-end distribution. We find that at the top 80% of the test score distribution, the teacher structure effect reaches 30 points to explain the school-type performance gap in mathematics. This implies that teacher characteristics exert a more substantial influence on student performance among high-achieving students. On the other hand, for Spanish, we find that the effect of teachers is more important on the lower part of the distribution but much smaller in magnitude. Additionally, by analyzing each variable included in teacher characteristics, we document that the most relevant are their own performance in the college admissions test and having an academic degree in education.

Finally, we find evidence of complementarity between students' prior performance and teachers' characteristics by analyzing teacher productivity differences on the test score distribution (residualized from students' pre-high-school test score performance). These insights are of great value to policymakers, educators, and researchers, offering valuable guidance in promoting educational equity and enhancing the quality of education for all students, regardless of their type of school.

To the best of our knowledge, this is the first paper to explore the transmission of teacher-student performance in the context of college admission tests in Chile. In a previous effort, Contreras (2002) studies the impact of the type of school on the college admissions test score; however, the author does not include teacher characteristics in the analysis. Other studies conducted in Chile have analyzed the impact of teacher characteristics on lower-stakes exams (Canales and Maldonado 2018; Toledo Román and Valenzuela 2015; Barrios Fernández and Riudavets 2021). In addition to looking at a higher-stake exam, our paper also includes teachers' own performance on the college admissions test as a relevant driver.

The remainder of the paper is organized as follows. Section 2 describes the institutional background of the educational system in Chile. Section 3 summarizes the previous literature, Section 4 describes the data, Section 5 performs an exploratory analysis of the main factors that determine students' performance in the college admissions test, Section 6 presents the methodology to decompose the achievement gap, Section 7 presents the results, and Section 8 concludes.

2 Institutional Background

The Chilean educational system is divided into eight years of primary education and four years of secondary education. There are three types of schools: public schools, funded and administered by the government; voucher schools, which receive partial funding from the government through a voucher system and are administered by the private sector; and private fee-paying schools, funded and administered by the private sector. Regarding the distribution of students, approximately 40%

Public schools are more likely to serve disadvantaged student populations than private voucher schools (Elacqua 2012). At the same time, voucher schools have been shown to attract better teachers than public schools (Behrman et al. 2016). Voucher schools face fewer regulations in their hiring system than public schools and have greater autonomy in their use of resources and the development of the academic curriculum. However, while within each school, the socioeconomic level of students tends to be highly homogeneous, voucher schools serve families from a broad socioeconomic spectrum, and approximately 80% of its enrollment is concentrated in lower, lower-middle, and middle socioeconomic groups.

Throughout their school years, students at specific grade levels participate in SIMCE tests (Sistema de Medición de la Calidad de la Educación), standardized assessments administered in Chile to assess the quality of education and school performance. These assessments are conducted nationally and focus on topics relevant to the respective grade levels, including mathematics and reading comprehension.

Successful completion of secondary education is a prerequisite for admission to higher education institutions in Chile. Most higher education institutions select their students using a centralized deferred acceptance admission system that only considers the performance of students in secondary education and a standardized national university entrance exam (PSU).[2] The PSU can be completed in the last year of secondary education (12th grade) at the end of the academic year. The PSU consists of two mandatory sections, Mathematics and Language and Communication (Spanish), and at least one of the other sections, Scientific Reasoning or History, Geography, and Social Sciences. Some private universities do not participate in the centralized system and have their own admission tests or criteria that may differ from the PSU.

Students seeking admission to public universities in Chile use an online platform to apply to multiple universities and programs simultaneously. After registering, students can browse the available programs and universities. They select the specific programs they wish to apply and rank their program choices in order of preference. Students are admitted to programs based on their selection rank and the number of slots available.

The Chilean education system has faced challenges related to inequality in access, educational quality, and funding disparities between public, voucher, and private schools. Recent educational reforms have aimed to address these issues, including efforts to increase funding for disadvantaged students. However, these disparities remain a complex and persistent issue in Chile's education system. For example, in 2016, a centralized school admission system was introduced in an effort to make the admission process more equitable. However, recent evidence has not found a significant effect of this policy on reducing school segregation (Kutscher, Nath, and Urzúa 2023).

2. There are some few exceptions that include Special Admissions, which are reserved for students who meet specific criteria, such as athletes, indigenous students, or students with disabilities, and Admission by Merit, reserved for students with exceptional academic achievements or talents in specific fields.

3 Literature Review

There is a long-standing literature documenting how the quality of teaching significantly impacts students' academic performance (Hanushek et al., 2007; Chetty et al., 2014). The evidence suggests that teachers are among the most influential factors in explaining student achievement (Hanushek 2011). In particular, several studies have shown that a one standard deviation improvement in teacher quality leads to a roughly 0.1 standard deviation increase in student test scores (Rockoff 2004; Rivkin, Hanushek, and Kain 2005; Aaronson, Barrow, and Sander 2007).

The impact of teachers extends beyond academic performance. Research by Jackson (2018) and Petek and Pope (2023) reveals that teachers also influence non-test score behaviors, such as absences and suspensions. These dimensions of teacher quality have been found to have a lasting impact on students' long-term outcomes. In a different context, Chetty, Friedman, and Rockoff (2014b) shows that students assigned to better teachers are more likely to attend college and earn higher salaries.

Although there is broad consensus on the importance of teacher quality, accounting for it remains a challenging task, and studies differ on the importance of specific teacher factors in enhancing students' outcomes. In recent years, the adoption of Value-Added Models (VAMs) has become prevalent in educational research. For example, for the United States, Chetty, Friedman, and Rockoff (2014a) estimate that a standard deviation improvement in teacher value-added increases normalized test scores by 0.14 and 0.1 standard deviations in math and English, respectively. However, a common criticism of VAMs is their limited focus, namely, identifying overall teacher contributions to learning but providing little information on which teacher characteristics contribute more to improving student outcomes (Wei et al. 2012). We aim to contribute to this issue by analyzing the impact of different dimensions of teacher characteristics on high-stakes test score performance.

Most studies on the impact of teacher quality have focused on the US context. Nonetheless, a handful of studies have focused on the case of Chile. For example, Canales and Maldonado (2018) finds that teacher quality significantly affects eighth-grade standardized test scores, especially in math. They found no significant effect of teacher credentials but showed that the impact of teachers increases with professional experience. Similarly, Toledo Román and Valenzuela (2015) show that attributes such as short-term specific professional training and better curriculum coverage positively impact the performance of fourth-grade students. Barrios Fernández and Riudavets (2021) conduct teachers' VAMs and find that higher-quality teachers positively affect student test scores, high school graduation, higher education attendance, and the type of higher education institutions attended.

In a recent study, García-Echalar, Poblete, and Rau (2023) used VAMs to investigate the impact of teachers on gender gaps in standardized test scores. Their results reveal that, in general, teachers do not account for the existing math or Spanish score gaps between the genders. Interestingly, their research uncovers variations dependent on school type, with teacher value-added measures mitigating gender gaps in voucher schools but showing no such effect in public schools. This also motivates us to examine the school type's impact in our context.

In this paper, we take one step further and analyze whether teacher quality can explain the performance gap observed by different types of schools in Chile. Previous studies have examined the test achievement gap in various types of schools in primary and secondary education in Chile, using standardized test scores for students in the 4th, 8th, or 10th grades. For example, Bellei (2005) explored the relationship between school type and student performance in the 4th and 10th grades. Their findings indicate that, once accounting for the sorting of students due to selective admission processes and the exclusion of retained students, private schools are not more effective than public schools and may, in fact, be less effective. Furthermore, Mizala and Romaguera (2000) analyzed the performance gap in the SIMCE test scores. Their research revealed that the test score gap between vouchers and public schools disappears when controlling for family socioeconomic characteristics.

It is pertinent to investigate the college admissions test results, as they represent a high-stakes assessment in the educational context. Consistent with this, Contreras (2002) explores the influence of school type on college admissions tests in conjunction with other SES variables. The findings reveal that the school's effect on student performance in college admissions tests is notably substantial and statistically significant, even after controlling for parental education levels. In this paper, we exploit a much richer dataset that allows us to control for a more comprehensive set of teacher variables, including the teachers' own performance in college admission test assessments. Recent evidence by Neilson et al. (2022) shows a positive and concave relationship between pre-college academic achievement and later teacher productivity. Their evidence suggests that college entrance exams could be useful to screen out or recruit students entering teacher colleges. This underscores the potential role of including teachers' standardized college admission tests as a proxy for their productivity.

4 Data

We combine data from multiple sources to explore the elements influencing students' PSU performance. We use an unchanging individual masked identifier that enables us to establish connections between students, their teachers, and the teachers' historical performance and educational decisions, as it remains consistent across different administrative datasets. This section elaborates on what we can recover from each data set and the sample restriction we must impose to define our sample.

We have access to DEMRE (*Departamento de Evaluación, Medición y Registro Educacional*) data on the national college admission test results for all students taking the test between 2006 and 2021. We use these data to identify teachers' performance in this test before entering higher education and assess students' performance from the cohorts between 2013 and 2021. As some students retake the PSU, we only keep their first PSU. It is important to note that not all graduating students take the PSU, as it is not mandatory.

We merge eighth-grade SIMCE records for each student in math and Spanish tests and information on their gender and their mothers' highest educational degree attained (high school, technical, professional, post-graduate). However, due to the design of the SIMCE assessments, there are certain years when eighth-graders were not tested. Only six of the nine possible student cohorts with PSU score data underwent an eighth-grade SIMCE examination. Our methodology relies on a value-added specification, so this information is important, as it considers the entire history of students' input before high school. Therefore, we limit our analysis to those students who took the SIMCE in eighth grade and subsequently took the PSU in their senior year. We further restrict the sample to students for whom we observe the school attended during each of the high school years. This enables a precise assignment of teachers according to each student's classroom.

We retrieve each teacher's subject information for each classroom, grade, and year from administrative records. This information enables us to identify whether a teacher is responsible for teaching Spanish or math to the students in our sample. These records encompass additional attributes, such as gender, age, years of teaching experience, and whether they have an educational degree. Additionally, we use DEMRE datasets to access information related to teachers' PSU performance, their ranking in college applications, and the institution they selected.

The nature of this context is multidimensional since each student can potentially have multiple teachers for various subjects throughout their high school years. Conversely, each teacher instructs several students in each classroom each year. As this paper's objective centers on examining disparities in student PSU score distributions, we aggregate teacher characteristics throughout each student's secondary education. We do this by sequentially averaging the characteristics of teachers of the corresponding subject at the classroom level for each grade. If no teacher information is available for a classroom, we impute information in two ways. First, if available, we impute information on the average characteristics of the rest of the classrooms in the same cohort, grade, and school. If information is still missing, we impute the school average. This approach leaves each student as the primary unit of observation.

It is also important to acknowledge some sample-bound limitations. First, our analysis will be limited to exploring the impact of young teachers since we only have PSU data from 2006 onward. Second, as discussed previously, not all students have a young teacher, so we have to impute the information on the average teacher characteristics at the cohort-school level. Third, to ensure comparability in the results, we focus on students in the regular education system, leaving out students attending special education due to a disability or incarceration and those attending night school. Consequently, we end with a final sample of 428,973 observations for math and 415,315 for Spanish, with 307,169 students appearing in both subject samples.

4.1 Descriptive statistics

Table 1 presents the summary statistics of all student characteristics (Panel A) and average teacher characteristics (Panel B) for the sample of students considered in the analysis. Columns (1) to (4) refer to characteristics of public school students in the sample, while columns (5) to (8) refer to characteristics of voucher school students in the sample, with columns (1) and (5) showing average values for the math sample in each type of school, and (3) and (7) for Spanish.

The first row in Panel A presents the statistics for the main dependent variable, the PSU score. For math, we see in column (1) that public school students score on average 484.7, while column (5) shows that voucher school students score higher, 508. The pattern is similar when comparing results for Spanish: 477.9 vs. 506.1. The difference between these two values is the main focus of this paper, the unconditional PSU gap, rounding 23.36 points for math and 28.15 points for Spanish, representing a difference of about 0.2-0.3 standard deviations. To further put into perspective how large this gap is, consider that the average difference in year-to-year changes in cut-offs for admission into undergraduate programs is about 15 points, and the median of this difference is only 10 points. Figure 1 presents the distribution of PSU scores for students in the sample in math and Spanish, showing that the gap between schools is present not only for the mean value but for most of the distribution.

The following rows in Panel A show some additional characteristics of the sample. On average, public school students outperformed voucher school students in the eighth-grade knowledge test SIMCE by 0.22 standard deviations in math and 0.12 standard deviations in Spanish. These differences can be easily seen in Figure 2, where the difference in the distributions for math is much more severe than it is for Spanish. The lagged fraction taking each subject-specific PSU in each type of school also differs, with only 72% of students in public schools taking the tests, while the proportion in voucher schools reaches 78%. Other characteristics appear to be much more balanced between the types of schools, with around 55% of the test-taker students being female, and with 7-10% of mothers holding a technical degree, 28-30% holding a professional degree, and 3-5% of them holding a post-graduate degree and the rest, 55-62% holding at most a high school degree.

From Table 1, Panel B, we also learn about the differences in the average characteristics of the teachers in each type of school. We see that voucher school teachers, especially Spanish teachers, score much higher than public school teachers in the PSU of the subject they teach, as presented in Figure 3, but, nevertheless, public school teachers had higher grade point averages when graduating from high school than voucher school teachers, although on a much smaller scale than their PSU differences. Public school teachers are two to three percentage points less likely to hold a degree from a highly selective institution but slightly more likely to hold an education degree. Interestingly, it is more common for Spanish teachers to list education as a top 3 choice in their college application ranking than it is for math teachers, with teachers in public schools having a lower tendency to list it than voucher school teachers. The sample of teachers in this exercise is, of course, conditional on having taken the PSU. Since we have access to PSU information starting from the 2004 application process, but it takes some years for teachers to graduate and show up in the classrooms, the teachers we find are mostly young teachers beginning their professional careers, corresponding to only a portion of the teachers in the system. They are, on average, around 31 years old and have only three to four years of teaching experience in both types of schools, with half of the math teachers and three-quarters of the Spanish teachers being female.

5 Exploratory analysis: Predicting Academic Success

In this section, we explore the factors that influence student achievement in college admission tests, with a particular emphasis on the impact of teachers and schools. It is a well-established fact that socioeconomic characteristics, schools, and particularly teachers, are strong predictors of students' performance. To confirm this, as a first step, we estimate the following value-added model separately for each subject, math and Spanish, to predict college admission performance:

$$Y_{i,s,t} = \beta_0 + \beta_1 Voucher_{i,s,t} + \beta_2 \mathbf{SES_{i,s,t}} + \beta_3 \mathbf{Teacher_{i,s,t}} + \beta_4 SIMCE^{8b}_{i,s,t} + \beta_5 PSU^{takeup}_{s,t-4} + \gamma_t + \varepsilon_{ist} \tag{1}$$

where i denotes a student, s a school, and t a year. The outcome of interest $Y_{i,s,t}$ corresponds to the PSU score of the student i' in the year t graduating from high school s. $Voucher_{i,s,t}$ is an indicator variable taking value one if the students attended a school is a voucher school in high school, zero if it is public.[3] This variable captures any gap between comparable students in each type of school. $SES_{i,s,t}$ is a vector that encompasses variables that proxy the socioeconomic status of the student, specifically the education of the mother of the student (no higher education degree, technical tertiary degree, university degree, or graduate degree). $Teacher_{i,s,t}$ represents a vector that encompasses the mean characteristics of the teacher observed throughout the high school years of a student. This vector incorporates several factors, including the average performance of teachers in the PSU in the subject they teach, their mean high school GPA (measured on the PSU scale), the proportion of teachers who hold an education degree, the fraction of teachers who ranked education among their top three choices in college applications, and the proportion of teachers who attended a selective university.[4]. Additionally, this variable includes average teacher experience, age, and the proportion of female teachers. The variable $SIMCE^{8b}_{i,s,t}$ captures the students' performance on the eighth-grade SIMCE test of the corresponding subject. This allows us to interpret the results as a value-added specification, where SIMCE is taken to be a sufficient statistic for the inputs of the students prior to high school (Todd and Wolpin 2003). Lastly, to control for possible selection on the PSU takeup across schools, we control for the proportion of students in the school who took the PSU test four years earlier, which is captured by $PSU^{takeup}_{s,t-4}$. We calculated it lagged to reduce concerns about potential endogeneity issues.[5] We also include in the estimations the gender of the student, year-fixed effects, γ_t, to capture aggregate shocks to PSU results at the national level.

The fact that we can incorporate the teacher's own performance in the college admissions test and their application preferences when they apply to college is a novel contribution to the existing literature. To the best of our knowledge, we are the first to attempt to study the relationship between student performance in college admissions tests and teachers' performance in the very same test.

5.1 Regression Results

Table 2 presents the results predicting performance in the math PSU. We start by looking at the impact of the type of high school attended by the students and gradually incorporate the rest of the relevant variables into the following columns. Column (1) presents the results of including only

3. Only 6.19% of students in our sample switched from a voucher to a public school, or vice-versa. For these students, we consider the school they graduated from.

4. The select universities considered are Pontificia Universidad Católica de Chile (PUC) and Universidad de Chile (UCH), which are the most selective institutions in the country (Bordón, Canals, and Mizala 2020)

5. As Table 1 indicates, approximately 70% of high-school students take the PSU. This figure might raise concerns about the impact of self-selection in the college admission test on our estimates. Since we examine the impact of school type, we cannot use school-fixed effects to account for this potential source of bias. However, we interpret our lagged measure of school-specific propensity to produce test takers and the student's pre-high school performance as sufficient statistics accounting for this self-selection into testing.

the voucher-school indicator variable in the estimation and fixed effects of the college application year. The voucher coefficient indicates the average differences in math PSU scores between the two types of schools. It indicates that the expected conditional gap between school types is 23.3 PSU points. Column (2) indicates that the PSU gap drops to 22 points when we include controls for student background, such as gender and the maternal education indicator variables.

Column (3) includes teacher information, showing that some of them are strong predictors of student performance. The average teacher's own math PSU result is a powerful positive predictor of student math PSU performance, with a one standard deviation increase in teacher score associated with an increase in 9 points, about 0.1 standard deviations. Having a higher proportion of teachers with a formal education degree also has a positive statistically significant relationship with the PSU score, indicating that an increase of 10% in the fraction of teachers with an education degree is associated with an increase of 3.7 points in the PSU score. There also is a tension between average years of experience and average teacher's age, where they seem to cancel each other's impact, which might be due to the teacher sample being particularly young and the presence of collinearity and little variation on both of them. Having a higher fraction of teachers with a degree from a selective institution, selecting education as a top 3 choice in their college application ranking, nor the proportion of female teachers are relevant predictors after controlling for the previously mentioned variables.

Note that after controlling for SES characteristics of the students and a comprehensive set of teacher characteristics, the type of school attended in high school still plays an important role in explaining PSU performance. When we include the eighth-grade math SIMCE score in column (4), the school gap reduces to 9.4 points. The coefficient for SIMCE in eighth grade indicates that having one standard deviation higher math SIMCE score is associated with a 62-point increase in the predicted math PSU score, making it a substantial predictor of academic success. We interpret this result as indicating that the inputs the student received in primary education play a significant role in explaining performance differences between students, reducing the relevance of the type of high school attended. Since our interest lies in explaining PSU results, including previous scores is ideal for our setting, in which the contents of the test are mostly prepared during high school.

Finally, in column (5), we include a proxy for the propensity of the school of origin to have students taking the PSU. We find that by including this control, the gap is reduced to only 7.5 points. This is consistent with the fact that, as presented in Table 1, voucher schools are 6 percentage points more likely to have their students take the PSU. Note that, likewise, the coefficients for teacher characteristics shrink in size after including SIMCE results and lagged PSU takeup rates up to a third of what they were in column (3). By including past test scores, the estimate is of a value-added model, netting out from the effects of the components that were coming through their pre-high school education, also allowing us to control for within-school heterogeneity in student performance potential, which appears to be substantial from the considerable increase in the goodness-of-fit of the model.

We find similar results when we turn to Table 3. Column 1 shows an expected conditional voucher gap of 28.3 points, which drops to 25.6 points in column (3) when controlling for students' gender, socioeconomic characteristics, and teacher characteristics. The results are very similar in size and direction to those for math PSU. The average Spanish PSU of teachers strongly correlates with a student's Spanish PSU performance and the fraction of teachers with an educational degree. Column (4) indicates a large improvement in the fit of the model, showing that the relevance of the school type in high school decreases by 7 points when including students' eighth-grade Spanish SIMCE scores. Lastly, column (5) presents that the school gap further goes 2.5 points down by including the school PSU takeup proxy, with teacher characteristics estimates going down to about half the size they were in column (3).

These results impose that the coefficients for each explanatory variable need to be the same across both voucher and public schools, but that might not necessarily be the case if there are productivity differences in the use of any of those variables. Guided by the results in Tables 2 and 3, we apply the methodology developed by Firpo, Fortin, and Lemieux 2018 and Rios-Avila 2019 to analyze whether teacher quality can explain the performance gap observed by different types of schools. Hence, in the next section, we estimate the decomposition of the school type gap in the average PSU performance by the contribution of socioeconomic characteristics, teachers, own past performance, and propensity to take the PSU test, allowing more flexibility to explore the isolated effects coming from the different elements of the model.

6 Bridging the school-type gaps

The regression analysis in the previous section identifies many influential factors that explain the gap in PSU performance measured as the difference in the means of the scores of students in the two types of schools. We implement a classical Oaxaca-Blinder decomposition to separately estimate how much of the difference comes from the composition effect, i.e., the differences in covariates between the two groups, and how much comes from the structure effect, i.e., the estimated coefficients, which in this educational setting is akin to the productivity of the covariates. Then, we take the analysis one step further by implementing a RIF Oaxaca-Blinder Decomposition, which allows us to go beyond simple mean comparisons to consider gaps in other statistics independent of the decomposition's sequential order. We follow Firpo, Fortin, and

Lemieux 2018 to conduct the RIF Oaxaca-Blinder type decomposition analysis to explain the differences in PSU performance of students between schools for both the mean and quantiles, separating the differences in the distributions into composition and structure effects, and allowing us then to decompose each effect by the contribution of each covariate, combining the Recentered Influence Functions (RIF) analysis and the reweighted strategy proposed by DiNardo, Fortin, and Lemieux 1996, following the nomenclature as laid out in (Rios-Avila 2019).

Assume that there is a joint distribution that describes all relationships between PSU scores, Y, exogenous characteristics, X, and the categorical variable indicating the types of schools to be compared, T. Then, we can rewrite the PSU distribution conditional on school type as:

$$\begin{aligned} f^k_{Y,X}(y,x) &= f^k_{Y|X}(Y|X) f^k_X(X), & (2) \\ F^k_Y(y) &= \int F^k_{Y|X}(Y|X) dF^k_X(X), & (3) \end{aligned}$$

where k indicates whether the density is conditional on the type of school, $T = k$ with $k \in \{0,1\}$. Then, differences in any distributional statistic v can be calculated as:

$$\begin{aligned} \Delta v &= v_1 - v_0 \\ &= v(F^1_Y) - v(F^0_Y) \\ &= v(F^1_{Y|X}(Y|X) dF^1_X(X)) - v(F^0_{Y|X}(Y|X) dF^0_X(X)). & (4) \end{aligned}$$

From equation (4) it follows that the differences in statistics v can arise from differences in average characteristics $\left(dF^1_X(X) \neq dF^0_X(X))\right)$, or differences in coefficients $\left(F^1_{Y|X}(Y|X) \neq F^0_{Y|X}(Y|X)\right)$. To separately estimate how relevant the composition and structure effects are in separately explaining the school-type gap, it is needed a third statistic, a counterfactual one, that permits the consideration of step-wise variations:

$$v_c = v(F^c_Y) = v(F^0_{Y|X}(Y|X) dF^1_X(X)).$$

With this counterfactual statistic, we can decompose Δv in equation (4) as:

$$\Delta v = \underbrace{v_1 - v_c}_{\Delta v_s} + \underbrace{v_c - v_0}_{\Delta v_x},$$

where Δv_s denotes the structure effect and Δv_x represents the composition effect. However, v_c is, by definition, a counterfactual statistic and, therefore, not observable in the data. This represents an empirical challenge, which can be sorted out by approximating the relevant distribution as follows:

$$F^c_Y(y) = F^0_{Y|X}(Y|X) dF^1_X(X) \approx F^0_{Y|X}(Y|X) dF^0_X(X) \omega(X),$$

where the weights, $\omega(X)$, can be obtained applying Bayes rule:

$$\omega(X) = \left[\frac{1-P}{P}\right] \times \left[\frac{P(T=1|X)}{1-P(T=1|X)}\right],$$

where P is the proportion of students in school type $T = 1$, and $P(T = 1|X)$ is the conditional probability that someone with characteristics X belongs to a school type $T = 1$. Thus, by re-weighting $dF_X^0(X)$, we can proxy for v_c.

We estimate $P(T = 1|X)$ in a logit model, using as the main explanatory variables the percentage of voucher schools in the municipality of residence of students, the total number of voucher schools in the municipality of residence of students, mother's education (less than high school, technical, professional or graduate degree), student-level eighth-grade SIMCE scores on both subjects, gender, the average experience of teachers in the municipality of residence, and average PSU scores of teachers in the municipality of residence. We also include fixed effects of the year and interactions between the proportion of voucher schools, the total number of voucher schools, SIMCE scores, and the average characteristics of teachers with the SIMCE scores of students and the level of education of the mother.[6] Thus, we have:

$$v_i = E\left(RIF\left(y_i; v(F_Y^i)\right)\right) = \bar{X}^{i\prime}\hat{\beta}_i \qquad \text{for } i \in \{0,1,c\}.$$

We can then decompose the gaps in the PSU scores between the two types of schools, public and voucher, as follows:

$$\Delta v = \underbrace{\bar{X}^{1\prime}\left(\hat{\beta}_1 - \hat{\beta}_c\right)}_{\Delta v_s^p} + \underbrace{\left(\bar{X}^1 - \bar{X}^c\right)'\hat{\beta}_c}_{\Delta v_s^e} + \underbrace{\left(\bar{X}^c - \bar{X}^0\right)'\hat{\beta}_0}_{\Delta v_x^p} + \underbrace{\bar{X}^{c\prime}\left(\hat{\beta}_c - \hat{\beta}_0\right)}_{\Delta v_x^e}, \tag{5}$$

where the structure effect is further divided in pure structure $\left(\Delta v_s^p\right)$ and a reweighting error (Δv_s^e). Likewise, the composition effect is separated into the pure composition effect $\left(\Delta v_x^p\right)$ and a specification error (Δv_x^e).

The intuition behind pure composition effects, Δv_x^p, is to capture differences in PSU performance between groups that can only be explained by the fact that the two groups are different. For example, we know that voucher school students score higher in 8th-grade knowledge exams than public school students. Therefore, we would expect them to also have an advantage on subsequent college admission test results over the other group of students. This kind of difference between groups would be isolated in the composition effect. For its part, pure

6. This richer specification with many interaction terms is needed to improve the fit in the reweighting process. See also Lemieux 2002.

structure effects, $\Delta \nu_s^p$, would indicate differences in PSU scores that are due to factors that are more productive for one type of school than the other, leading to better PSU results under the same levels of factors. If, for example, keeping all student characteristics constant, having a more experienced teacher is more advantageous (productive, as measured in PSU score points) for voucher school students, this effect would be captured in the structure effect.

The two additional estimates from the RIF OB decomposition are the reweighting and the specification error, $\Delta \nu_s^e$ and $\Delta \nu_x^e$. The reweighting error comes from the selection of the variables and interaction terms included to compute the counterfactual statistic by estimating $P(T=1|X)$, and should go to zero in large samples. Of course, there is a tension between a higher $Pseudo-R^2$, a common support, and a perfect prediction, which is undesirable (Firpo, Fortin, and Lemieux 2018). Lastly, the specification error comes from departures from linearity in the conditional expectation and from the fact that $F_Y^c(y)$ is an approximation, so we should expect this error to be different from zero.[7]

In the next section, we estimate the Oaxaca-Blinder classical and RIF decomposition to explain the gap separately in the math and Spanish results, following Firpo, Fortin, and Lemieux 2018. Considering the evidence presented in Tables 2 and 3, we exclude from teacher characteristics the non-significant components (selective education dummy, education as top choice dummy, and teacher's gender), with the objective of minimizing the noise in the decompositions. We implement this routine for two different distribution moments: mean and quantile gaps.

7 Main results

We first implement a conventional Oaxaca-Blinder decomposition of the form:

$$E[PSU|X,\text{Voucher}] - E[PSU|X,\text{Public}] = \underbrace{\bar{X}^V\left(\hat{\beta}_V - \hat{\beta}_P\right)}_{\text{Unexplained}} + \underbrace{\left(\bar{X}^V - \bar{X}^P\right)\hat{\beta}_P}_{\text{Explained}},$$

where $\bar{X}^k$ denotes a vector containing the averages of the independent variables for students enrolled in type k's schools, and $\hat{\beta}_k$ is the associated vector of point estimates obtained from a linear regression model ($k \in \{Public, Voucher\}$). As it is standard in the literature, the school gap attributed to the means is denoted "explained", whereas the part attributed to the coefficient is the "unexplained" gap. To implement this decomposition, we employ the set of controls in equation

7. As Firpo, Fortin, and Lemieux 2018 point out, how large the error should be is an open empirical question.

(1). Notice that under this specification, we obtained small point estimates for the average Math and Spanish PSU gaps (see Column 5 in Tables 2 and 3, respectively).

Table 4 presents our results. Panel A displays the overall gaps. Consistent with the summary statistics, voucher schools have unconditional average advantages in Math (Column 1) and Spanish (column 2) PSU scores. However, 90% and 76% of these gaps, respectively, are explained by the average differences in the observed characteristics. Panel B shows that the most important contributors are eighth-grade test scores and school-level historic PSU takeup. High school teachers' characteristics, a compound of the different variables in this category, play a relatively minor role in both subjects.

The analysis of the contributors to the unexplained gaps delivers a different story. Panel C of Table 4 suggests that the coefficients associated with the teacher's characteristics contribute more than 31 points to the Math PSU score gap (favoring voucher schools), the largest contributing factor. For Spanish, the impact is much smaller in absolute size and not statistically significant. Overall, these results suggest that equipped with the same inputs, public and voucher schools produce different outcomes, suggesting differences in productivity levels across school types. We come back to this point later.

Table 5 presents the results of the general version of the Oaxaca-Blinder decomposition, which includes the reweighting scheme (expression (5)). Columns 1 and 2 display the results for math and Spanish, respectively. Panel A shows that observed characteristics (Composition) explain a larger proportion of the total gaps than the parameters (Structure). The analysis of the Composition effects (Panel B) indicates that 8th-grade test score is the most important contributor, followed by the School's PSU takeup (lagged). Teachers' characteristics contribute with less than one PSU point to closing the gap. However, as reported in Table 4, the coefficients associated with these variables play an important role in closing the average gaps, particularly for Math. Panel C indicates that more than 22 points of the Math gap can be attributed to this "structure" component. The school PSU takeup rate is also an important contributor, confirming the role of selection in PSU discussed above. SES characteristics and the eighth-grade test scores contribute marginally to this component.

These results confirm that pre-high school test scores and the school's (predetermined) college admission test takeup emerge as the most important differences in characteristics explaining the average PSU gap between public and voucher schools. After controlling for these variables, the gaps reduce and even disappear. This limits the extent to which schools can modify and adjust input (e.g., teacher characteristics) to reduce the gaps in a specific cohort. Now, since the differences in coefficients in Equation (5) can be interpreted as proxies for the differential productivity levels of schools, our findings also suggest that, at equal input, voucher schools are better at producing higher PSU scores. This represents a major challenge for public policies, and it is consistent with the long-standing evidence documenting the voucher schools' unconditional test scores' advantages.

7.1 Beyond averages: Quantile gaps

Decomposing the mean differences in PSU scores between public and voucher schools is informative of the factors driving these gaps and the effectiveness of public initiatives to close them. However, this approach does not reveal the factors that affect students at different distribution points. For example, for low-performance students, the drivers of gaps between public and voucher schools could differ from those affecting students in the middle or at the top of the distributions. To examine this, we implement the decomposition introduced in Section 6, which expresses the differences in any distributional statistic as the sum of the structure and composition effects.

Figures 4 and 5 represent the results for Math and Spanish, respectively. Given the similarities in their messages, we focus mostly on Math and discuss any disparities between the two subjects.

Panel A of Figure 4 presents the overall difference in Math PSU between public and voucher schools, decomposing it into composition and structure effects in each quantile using the reweighting procedure described in equation (4). The estimated overall difference (red line) is more or less stable across the distributions of the PSU scores, decreasing slightly as we move up across the quantiles. The range lies in the 20 to 35 interval, with an average of 30 points. This stability suggests that the distribution of student-level scores of voucher schools is shifted to the right relative to public schools. This is consistent with the evidence in Figure 1, which displays the distributions. The positive and increasing composition effects (blue dotted line) indicate that this component increasingly explains the gaps, with the observed characteristics increasingly favoring voucher schools as one moves up in the distributions. We come back to this point below. Finally, the structure effects, depicted by the dashed green line in Panel A, partially compensate for the composition effects, showing a declining slope toward the highest quantiles.

Panel B of the same figure presents the contribution of the different sets of factors to the overall gaps between school types by quantiles. While there is almost no role for socioeconomic characteristics in explaining the gaps, teachers are the main driver in expanding them, particularly in the upper half of the distribution. Note that this is not the case for Spanish, which we discuss later in this section. Additionally, heterogeneity from the pre-high-school test scores (SIMCE) explains between one-third and two-thirds of the gaps across the whole distribution. Finally, consistent with the findings of Tables 2 and 3, accounting for PSU enrollment reduces the advantage of voucher schools, making the gap negative as we move up in the distribution.

Panels C to E complement the previous results and provide further insight into the relative magnitude of the different effects. Specifically, Panel C shows that most of the composition effects come from the pure explained component; meanwhile, Panel E shows that it is not the case for structure effects, where pure explained and residual effects mostly net out each other. Panels D and F show that the importance of pre-high-school test scores comes from the composition instead of structure effects while confirming that the school-type-specific coefficients (structure effect) of teachers' characteristics and predetermined PSU takeup rates are important drivers of disparities. It is interesting to observe that the contribution of teachers to the structure effect is somewhat different between math and Spanish (see Panel F of Figure and 5). For math, we see that this component is particularly important in explaining the gap in the upper half of the distribution, with voucher schools being relatively more productive. For Spanish, we observe that teachers are more important in explaining the gap in the lower half of the distribution, with voucher schools being relatively more productive. In the upper part of the distribution, the pattern is much more noisy.

7.2 Complementarities

Figure 2 shows that students from public schools perform worse on the SIMCE test than students from voucher schools. This fact holds for both subjects, although the gap is larger for math. Now, we explore how these disparities interact with teacher characteristics. Specifically, our analysis only assesses the impact of high school teachers and examines whether they can improve students' PSU scores independently of their previous academic performance. Whether this is true or not is what we try to analyze in this subsection.

We start by analyzing the point estimates from Equation 1 when introducing two interaction terms to the model. The first is the interaction between students' performance on the eighth-grade SIMCE test and their average teachers' performance on the college admissions test. We also control for the interaction between students' performance on the eighth-grade SIMCE test, their teachers' performance on the college admissions test, and the voucher school indicator variable. Table 6 reports these results.

Columns (1) and (4) present regression results, including all explanatory variables, in the estimations explaining the math and Spanish PSU scores, equivalent to the results in column (5) in Tables 2 and 3, respectively. Columns (2) and (5) present results from including the interaction term between the average teacher's PSU and the student's past SIMCE math and Spanish scores, respectively. We only observe positive and significant coefficients associated with the interaction term for math. This result suggests that the positive effect of having a teacher with a higher PSU score is amplified when the students have higher pre-high school test scores in the case of math

PSU scores. Lastly, columns (3) and (6) include an additional interaction term, multiplying the interaction term by the voucher indicator variable. We observe that the coefficient associated with the interaction between student and teacher performance and the voucher dummy is negative for both tests, although only statistically significant for math. Thus, the amplification of the teachers' PSU effect is smaller (and nearly nonexistent in Spanish) for students coming from voucher schools.

What would happen with the contribution of teachers to the performance gap between school types if we abstracted from the differences in the distribution of students' previous test scores and, therefore, from the complementarity between SIMCE's score and teachers' PSU scores? To answer this question, we re-estimate a RIF Oaxaca-Blinder decomposition of the distribution of students' PSU scores net of SIMCE, focusing only on the structure effect of teacher characteristics on this measure. To accomplish this, we create a new measure of student PSU performance orthogonal to SIMCE scores by residualizing the PSU scores by their own SIMCE test scores and using this residualized measure as a dependent variable. The results are graphically presented for quantiles of the distributions in Figure 6, with Panel A showing the results for math and Panel B for Spanish for both the original PSU score and the new residualized version.

Since the gap between vouchers and public schools might change once we decompose it using the residualized PSU measure, in Figure 6 we plot the contribution of the teacher's structure effect to the corresponding gap as a percentage of the total gap. We can observe that in both figures, the contribution of teachers is null overall and positive only at the top of the math test score distribution. This result suggests the existence of complementarities: once we net out the prior test score achievement, the contribution of the teachers to the gap disappears.

8 Conclusion

This paper explores the determinants of Chile's college admission test, with a focus on the role of teachers in explaining documented gaps in the nation's schooling system. The analysis exploits unique and rich matched teacher-student data gathered from multiple administrative information sources for the universe of test takers between 2013 and 2021.

We estimate a value-added specification and different cognitive achievement production function decompositions. Our findings show that teachers' performance on the college admission test (at age 17-18) and whether they hold a formal educational degree predict their students' success, along with the type of school attended in high school. We further examine whether teacher quality can account for the performance disparities observed across different types of schools. We find that productivity differences emerge as drivers of disparities across the distribution of test scores. Teacher-student interactions play an essential role.

Finally, by examining differences in teachers' productivity on the test score distribution residualized from students' previous test score performance, we find evidence of complementarities between students' prior performance and teachers' characteristics.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and student achievement in the Chicago public high schools." *Journal of labor Economics* 25 (1): 95–135.

Barrios Fernández, Andrés, and Marc Riudavets. 2021. "Teacher value-added and gender gaps in educational outcomes." *Available at SSRN 3856935.*

Behrman, Jere R, Michela M Tincani, Petra E Todd, and Kenneth I Wolpin. 2016. "Teacher quality in public and private schools under a voucher system: The case of Chile." *Journal of Labor Economics* 34 (2): 319–362.

Bellei, Cristian. 2005. "The private-public school controversy: The case of Chile." In *Conference on Mobilizing the Private Sector for Public Education, October,* vol. 5. 6. sn.

Bordón, Paola, Catalina Canals, and Alejandra Mizala. 2020. "The gender gap in college major choice in Chile." *Economics of Education Review* 77:102011.

Bravo, David, Sankar Mukhopadhyay, and Petra E Todd. 2010. "Effects of school reform on education and labor market performance: Evidence from Chile's universal voucher system." *Quantitative economics* 1 (1): 47–95.

Canales, Andrea, and Luis Maldonado. 2018. "Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics." *International journal of educational development* 60:33–50.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014a. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American economic review* 104 (9): 2593–2632.

———. 2014b. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American economic review* 104 (9): 2633–2679.

Contreras, Dante. 2002. "Vouchers, school choice and the access to higher education." *School Choice and the Access to Higher Education (June 2002).*

DiNardo, John, Nicole M Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64, no. 5 (September): 1001–1044.

Elacqua, Gregory. 2012. "The impact of school choice and public policy on segregation: Evidence from Chile." *International Journal of Educational Development* 32 (3): 444–453.

Firpo, Sergio P, Nicole M Fortin, and Thomas Lemieux. 2018. "Decomposing wage distributions using recentered influence function regressions." *Econometrics* 6 (2): 28.

García-Echalar, Andrés, Sebastián Poblete, and Tomás Rau. 2023. "Teacher Value-added and the Test Score Gender Gap." *Manuscript.*

Gilraine, Michael, and Nolan G Pope. 2021. *Making teaching last: Long-run value-added.* Technical report. National Bureau of Economic Research.

Hanushek, Eric A. 2011. "The economic value of higher teacher quality." *Economics of Education review* 30 (3): 466–479.

Heckman, James J. 2000. "Policies to foster human capital." *Research in economics* 54 (1): 3–56.

Iturra, Victor, and Mauricio Gallardo. 2022. "Schools, circumstances and inequality of opportunities in Chile." *International Journal of Educational Development* 95:102668.

Jackson, C Kirabo. 2018. "What do test scores miss? The importance of teacher effects on non–test score outcomes." *Journal of Political Economy* 126 (5): 2072–2107.

Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico. 2016. "THE EFFECTS OF SCHOOL SPENDING ON EDUCATIONAL AND ECONOMIC OUTCOMES: EVIDENCE FROM SCHOOL FINANCE REFORMS." *The Quarterly Journal of Economics* 131 (1): 157–218.

Kutscher, Macarena, Shanjukta Nath, and Sergio Urzúa. 2023. "Centralized admission systems and school segregation: Evidence from a national reform." *Journal of Public Economics* 221:104863.

Lemieux, Thomas. 2002. "Decomposing changes in wage distributions: a unified approach." *Canadian Journal of Economics/Revue canadienne d'économique* 35 (4): 646–688.

Mizala, Alejandra, and Pilar Romaguera. 2000. "School performance and choice: the Chilean experience." *Journal of human Resources,* 392–417.

Montaño, Sebastián, Catalina Morales, Cristina Riquelme, and Sergio Urzúa. 2023. "Do high-performing teachers produce high-performing students?" *Manuscript.*

Neilson, Christopher, Sebastian Gallegos, Franco Calle, and Mohit Karnani. 2022. "Screening and recruiting talent at teacher colleges using pre-college academic achievement."

Petek, Nathan, and Nolan G Pope. 2023. "The multidimensional impact of teachers on students." *Journal of Political Economy* 131 (4): 1057–1107.

Rios-Avila, Fernando. 2019. "Recentered influence functions in Stata: Methods for analyzing the determinants of poverty and inequality." *Levy Economics Institute, Working Paper* 927.

Rivkin, Steven G, Eric A Hanushek, and John F Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73 (2): 417–458.

Rockoff, Jonah E. 2004. "The impact of individual teachers on student achievement: Evidence from panel data." *American economic review* 94 (2): 247–252.

Todd, Petra E, and Kenneth I Wolpin. 2003. "On the specification and estimation of the production function for cognitive achievement." *The Economic Journal* 113 (485): F3–F33.

Toledo Román, Gabriela, and Juan Pablo Valenzuela. 2015. "Over-estimating the effects of teacher attributes on school performance in the Chilean education system." *Estudios de Economıa* 42 (1).

Wei, Hua, Tracey Hembry, Daniel L Murphy, and Yuanyuan McBride. 2012. *Value-added models in the evaluation of teacher effectiveness: A comparison of models and outcomes.*

Table 1: Public vs. Voucher Schools: Summary Statistics

	Public				Voucher			
	Math		Spanish		Math		Spanish	
	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Mean (5)	Std. Dev. (6)	Mean (7)	Std. Dev. (8)
Panel A. Student-level information								
PSU results	484.67	102.84	477.95	103.95	508.03	99.00	506.10	98.76
Simce 8th grade	-0.13	1.03	-0.08	1.01	0.09	0.97	0.04	0.98
Fraction taking PSU (lagged)	0.72	0.20	0.72	0.20	0.78	0.20	0.78	0.21
Female	0.55	0.50	0.57	0.49	0.55	0.50	0.55	0.50
Mothers with a technical degree	0.07	0.26	0.07	0.26	0.10	0.30	0.10	0.31
Mothers with a professional degree	0.30	0.46	0.30	0.46	0.28	0.45	0.28	0.45
Mothers with a postgraduate degree	0.04	0.18	0.03	0.18	0.05	0.22	0.05	0.22
Year Cohort	2,018.75	2.31	2,018.76	2.33	2,018.68	2.41	2,018.68	2.40
Panel B. Teacher's information								
Teacher PSU	-0.03	1.02	-0.10	1.02	0.06	0.98	0.11	0.97
Teacher GPA in PSU scale (NEM)	0.07	0.99	0.06	1.04	0.01	0.98	0.03	0.97
Graduated from PUC or UCH	0.08	0.25	0.06	0.23	0.10	0.28	0.09	0.28
Teacher has education degree	0.87	0.32	0.97	0.15	0.86	0.32	0.95	0.20
Education as top 3 choice in ranking	0.72	0.41	0.82	0.35	0.75	0.39	0.81	0.35
Years of teaching experience	3.55	3.39	3.80	3.35	3.29	2.95	3.28	2.71
Age	31.45	4.00	31.41	4.08	31.22	3.59	30.92	3.38
Female	0.53	0.46	0.75	0.40	0.52	0.45	0.74	0.40
Number of Observations	155,615		140,947		273,358		274,368	

Note: Panel A displays summary statistics of students in the sample, divided by the type of school they attend, public or voucher schools, and the subject of interest, math, and Spanish. Panel B displays similar summary statistics by type of school and subject but for the average characteristics of teachers teaching math or Spanish during high school to students in Panel A. The total number of students considered is 537,119, while for 307,169 of them, we have information on their performance and teachers' characteristics in both subjects.

Table 2: Public vs. Voucher Schools: Average Math PSU Gap
Regression Analysis

	Math PSU				
	(1)	(2)	(3)	(4)	(5)
Voucher	23.261***	21.951***	22.005***	9.442***	7.469***
	(4.773)	(4.709)	(4.326)	(2.244)	(2.165)
Teacher's Math PSU			9.058***	3.632***	3.390***
			(1.537)	(0.831)	(0.802)
Teacher's GPA			0.465	0.381	0.419
			(1.985)	(1.083)	(1.029)
Teacher has Selective Education			4.899	0.924	-0.098
			(5.317)	(2.980)	(2.889)
Teacher holds Education Degree			31.772***	13.351***	12.524***
			(3.313)	(1.652)	(1.553)
Education as Top 3 choice			-2.773	-1.610	-1.631
			(3.152)	(1.585)	(1.521)
Years of Teaching Experience			2.647***	1.141***	1.079***
			(0.550)	(0.285)	(0.271)
Teacher's Age			-1.726***	-0.826***	-0.810***
			(0.422)	(0.221)	(0.213)
Fraction Female Teachers			-0.662	-0.453	-0.549
			(3.474)	(1.833)	(1.735)
8th grade Simce (Math)				62.190***	61.145***
				(0.816)	(0.737)
School's PSU Takeup (lagged)					41.695***
					(3.039)
Number of Observations	428,973	428,973	428,973	428,973	428,973
R^2	0.025	0.032	0.054	0.413	0.418
Year FE	✓	✓	✓	✓	✓
SES Controls		✓	✓	✓	✓

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: The table presents the point estimates obtained from different versions of equation 1 using Math PSU as the dependent variable. The sample includes PSU takers covering the period 2013-2021. Year FE includes year-specific fixed effects for test-taking years. SES Controls include indicator variables for student gender and three indicator variables for maternal education categories: technical, undergraduate, and postgraduate degrees, with high school or less being the omitted category. Standard errors in parentheses clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Public vs. Voucher Schools: Average Spanish PSU Gap
Regression Analysis

	Spanish PSU				
	(1)	(2)	(3)	(4)	(5)
Voucher	28.305***	26.615***	25.666***	18.753***	16.064***
	(4.544)	(4.462)	(4.385)	(2.613)	(2.495)
Teacher's Spanish PSU			8.062***	4.781***	4.282***
			(1.532)	(0.930)	(0.878)
Teacher's GPA			-0.406	-0.341	-0.548
			(1.751)	(1.092)	(1.023)
Teacher has Selective Education			14.569	10.837**	9.124*
			(9.178)	(5.489)	(5.048)
Teacher holds Education Degree			24.545***	13.936***	12.958***
			(4.354)	(2.613)	(2.366)
Education as Top 3 choice			1.957	0.678	0.416
			(3.892)	(2.363)	(2.203)
Years of Teaching Experience			2.645***	1.267***	1.129***
			(0.489)	(0.295)	(0.282)
Teacher's Age			-1.556***	-0.705***	-0.630***
			(0.459)	(0.252)	(0.232)
Fraction Female Teachers			-2.005	-1.823	-1.300
			(2.887)	(1.791)	(1.701)
8th grade Simce (Spanish)				64.664***	63.753***
				(0.525)	(0.461)
School's PSU Takeup (lagged)					50.849***
					(3.602)
Number of Observations	415,315	415,315	415,315	415,315	415,315
R^2	0.019	0.029	0.043	0.433	0.441
Year FE	✓	✓	✓	✓	✓
SES Controls		✓	✓	✓	✓

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: The table presents the point estimates obtained from different versions of equation 1 using Spanish PSU as the dependent variable. The sample includes PSU takers covering the period 2013-2021. Year FE includes year-specific fixed effects for test-taking years. SES Controls include indicator variables for student gender and three indicator variables for maternal education categories: technical, undergraduate, and postgraduate degrees, with high school or less being the omitted category. Standard errors in parentheses clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Mean Decomposition
Oaxaca-Blinder without Reweighting

	(1) Math	(2) Spanish
A. Overall Gap		
(a) Average PSU in Voucher Schools	508.028***	506.102***
	(0.189)	(0.189)
(b) Average PSU in Public Schools	484.667***	477.950***
	(0.261)	(0.277)
Gap in favor of Voucher Schools ((a)-(b))	23.361***	28.153***
	(0.322)	(0.335)
Total Explained	16.065***	12.686***
	(0.215)	(0.242)
Total Unexplained	7.296***	15.467***
	(0.258)	(0.266)
B. Contributions to the Explained Gap		
Student's SES	0.545***	1.549***
	(0.035)	(0.050)
Teacher	0.168***	0.617***
	(0.031)	(0.062)
Student's Simce	12.815***	7.863***
	(0.194)	(0.215)
Lagged School PSU Takeup	2.537***	2.657***
	(0.063)	(0.069)
C. Contributions to the Unexplained Gap		
Student's SES	-1.764***	-2.526***
	(0.353)	(0.360)
Teacher	-8.062**	-35.140***
	(3.852)	(3.758)
Student's Simce	21.266***	-16.417***
	(1.536)	(1.395)
Lagged School PSU Takeup	-5.224***	-13.093***
	(0.966)	(0.996)
Intercept	1.079	82.642***
	(4.166)	(4.047)
Number of Observations	428,973	415,315

Note: The table presents the Oaxaca-Blinder decomposition for PSU scores in math and Spanish by type of school (public or voucher). The sample includes PSU takers covering the period 2013-2021. Standard errors in parenthesis clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Mean Decomposition
Oaxaca-Blinder with Reweighting

	Math	Spanish
A. Decomposition		
Total gap	23.361***	28.153***
	(0.31)	(0.317)
Composition	15.401***	12.237***
	(0.294)	(0.324)
Structure	8.282***	8.694***
	(0.3)	(0.308)
B. Contributions of Xs (composition)		
Student's SES	0.615***	1.639***
	(0.041)	(0.069)
Teacher's characteristics	-0.059	0.874***
	(0.036)	(0.057)
8th grade test score (Simce)	14.233***	9.101***
	(0.283)	(0.298)
School's PSU takeup (lagged)	0.611***	0.622***
	(0.039)	(0.042)
Specification error	-0.547*	7.476***
	(0.314)	(0.36)
C. Contributions of β s (structure)		
Student's SES	0.145	0.002
	(0.49)	(0.492)
Teacher's characteristics	9.937***	0.321
	(3.36)	(4.077)
8th grade test score (Simce)	0.051	-0.226***
	(0.032)	(0.029)
School's PSU takeup (lagged)	-10.17***	-19.241***
	(1.063)	(1.106)
Intercept	8.32**	27.838***
	(3.698)	(4.245)
Reweighting error	0.225	-0.254
	(0.314)	(0.248)
D. Total (composition + structure)		
Student's SES	0.759	1.64***
	(0.494)	(0.49)
Teacher's characteristics	9.878***	1.195
	(3.362)	(4.075)
8th grade test score (Simce)	14.284***	8.876***
	(0.285)	(0.29)
School's PSU takeup (lagged)	-9.559***	-18.618***
	(1.07)	(1.105)

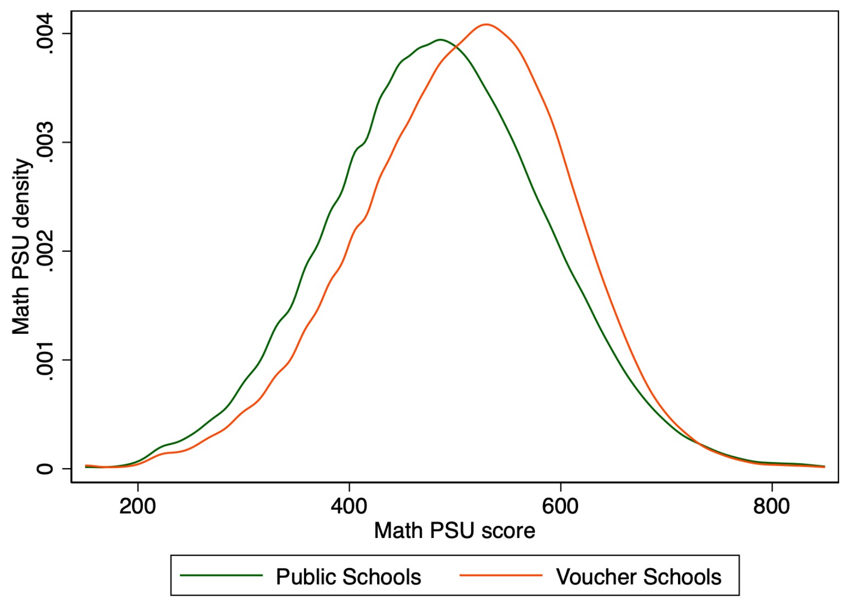
Note: The table presents the RIF Oaxaca-Blinder decomposition for the mean PSU score (math and Spanish) by type of school (public or voucher). The sample includes PSU takers covering the period 2013-2021. Bootstrapped standard errors over the entire procedure (100 replications) were used to compute the p-values and are presented in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Public vs. Voucher Schools: Student-teacher interactions as a mechanisms Regression Analysis Interaction

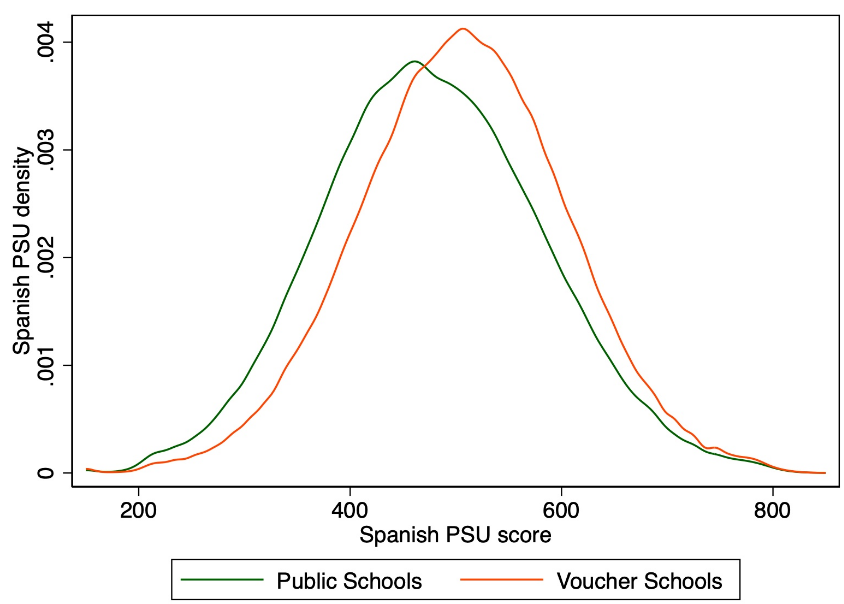
	Math PSU			Spanish PSU		
	(1)	(2)	(3)	(4)	(5)	(6)
Voucher	7.469***	7.647***	7.873***	16.064***	16.092***	16.200***
	(2.165)	(2.150)	(2.153)	(2.495)	(2.501)	(2.519)
Teacher's PSU	3.390***	3.442***	3.551***	4.282***	4.266***	4.299***
	(0.802)	(0.800)	(0.809)	(0.878)	(0.871)	(0.883)
Teacher's GPA	0.419	0.456	0.465	-0.548	-0.547	-0.546
	(1.029)	(1.033)	(1.039)	(1.023)	(1.023)	(1.026)
Teacher has Selective Education	-0.098	-0.511	-0.637	9.124*	9.099*	9.051*
	(2.889)	(2.819)	(2.805)	(5.048)	(5.060)	(5.088)
Education Degree	12.524***	12.337***	12.360***	12.958***	12.945***	12.935***
	(1.553)	(1.539)	(1.539)	(2.366)	(2.362)	(2.364)
Education as Top 3 choice	-1.631	-1.624	-1.611	0.416	0.421	0.446
	(1.521)	(1.512)	(1.511)	(2.203)	(2.204)	(2.201)
Years of Teaching Experience	1.079***	1.072***	1.074***	1.129***	1.131***	1.133***
	(0.271)	(0.273)	(0.272)	(0.282)	(0.281)	(0.281)
Teacher's Age	-0.810***	-0.797***	-0.800***	-0.630***	-0.632***	-0.634***
	(0.213)	(0.212)	(0.212)	(0.232)	(0.231)	(0.230)
Fraction Female Teachers	-0.549	-0.620	-0.690	-1.300	-1.300	-1.308
	(1.735)	(1.755)	(1.758)	(1.701)	(1.701)	(1.702)
Student's Simce	61.145***	61.106***	61.127***	63.753***	63.753***	63.812***
	(0.737)	(0.725)	(0.722)	(0.461)	(0.461)	(0.455)
School's PSU Takeup (lagged)	41.695***	41.723***	41.574***	50.849***	50.845***	50.807***
	(3.039)	(3.026)	(3.020)	(3.602)	(3.603)	(3.607)
Teacher PSU × Student SIMCE		2.780***	4.070***		0.373	1.264
		(0.415)	(0.719)		(0.365)	(0.795)
Teacher PSU × Student SIMCE × Voucher			-2.169**			-1.454
			(0.860)			(0.889)
Number of Observations	428973	428973	428973	415315	415315	415315
R^2	0.418	0.419	0.419	0.441	0.441	0.441
Year FE	✓	✓	✓	✓	✓	✓
SES Controls	✓	✓	✓	✓	✓	✓

Note: The table presents the point estimates obtained from different versions of equation 1 using Spanish PSU as the dependent variable. The sample includes PSU takers covering the period 2013-2021. Year FE includes year-specific fixed effects for test-taking years. SES Controls include indicator variables for student gender and three indicator variables for maternal education categories: technical, undergraduate, and postgraduate degrees, with high school or less being the omitted category. Standard errors in parentheses clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Distribution of student-level PSU scores by School Type



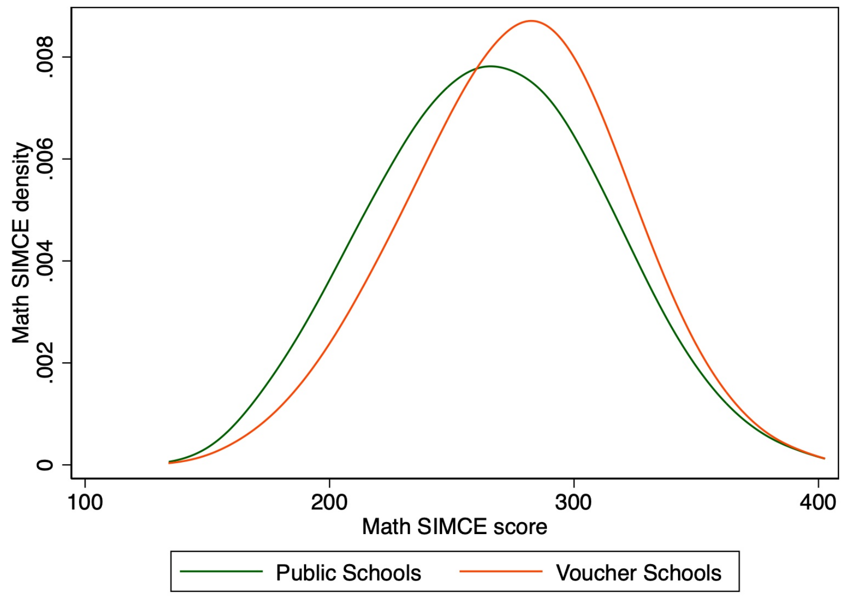
(a) Math PSU



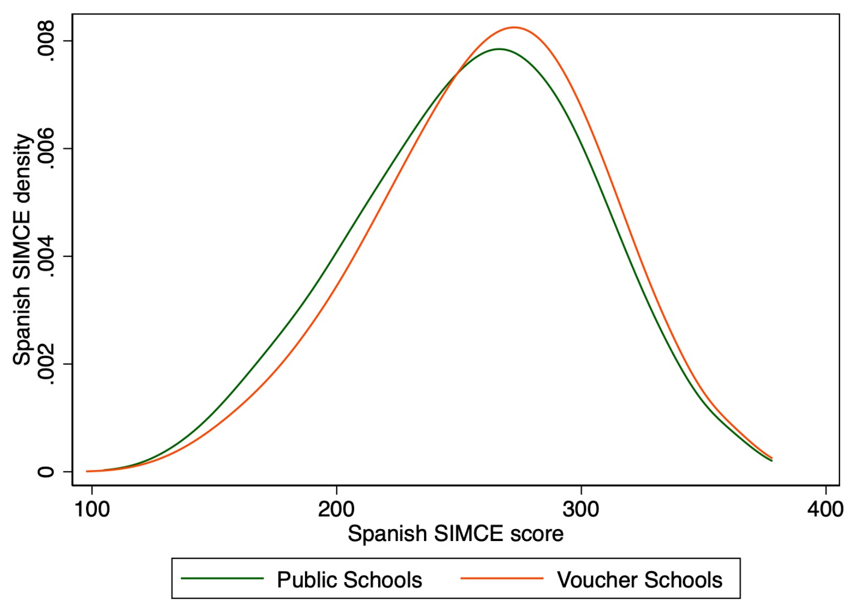
(b) Spanish PSU

Note: Panel A (B) displays the distribution of Math (Spanish) PSU scores computed from our sample of 428,973 (415,315) test takers between the years 2013 and 2021.

Figure 2: Distribution of student-level SIMCE scores by School Type



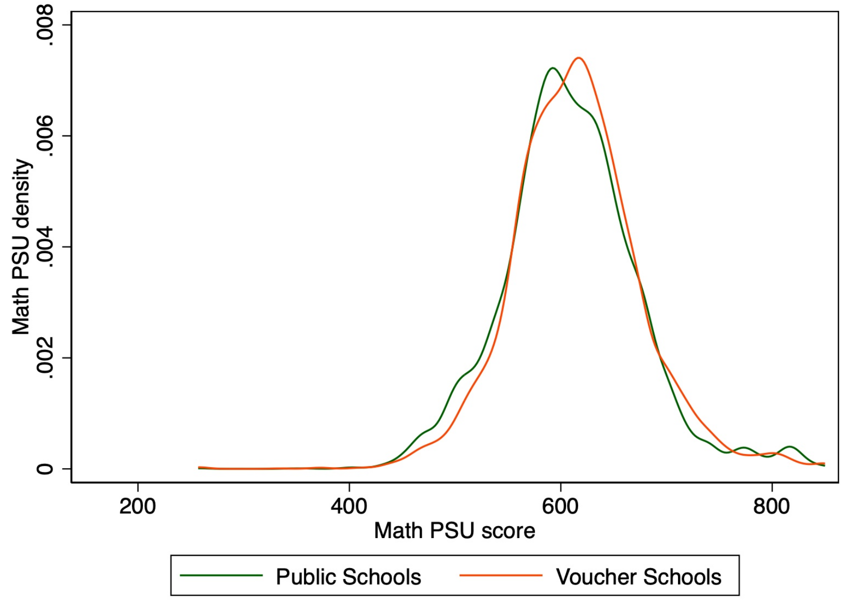
(a) Math SIMCE



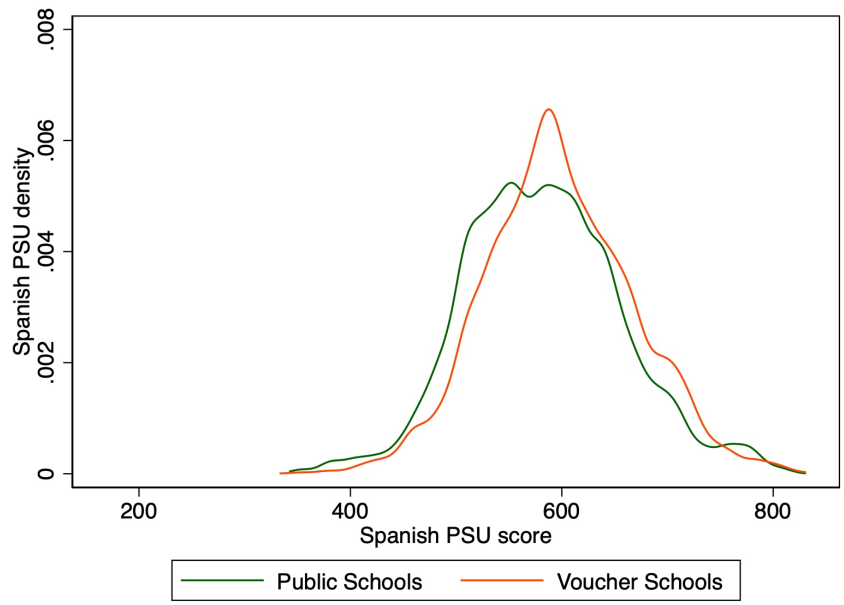
(b) Spanish SIMCE

Note: Subfigure a (b) displays the distribution of PSU scores at the student level for math (Spanish) computed from our sample of 428,973 (415,315) test takers between the years 2013 and 2021.

Figure 3: Distribution of Average Teachers' PSU scores by School Type



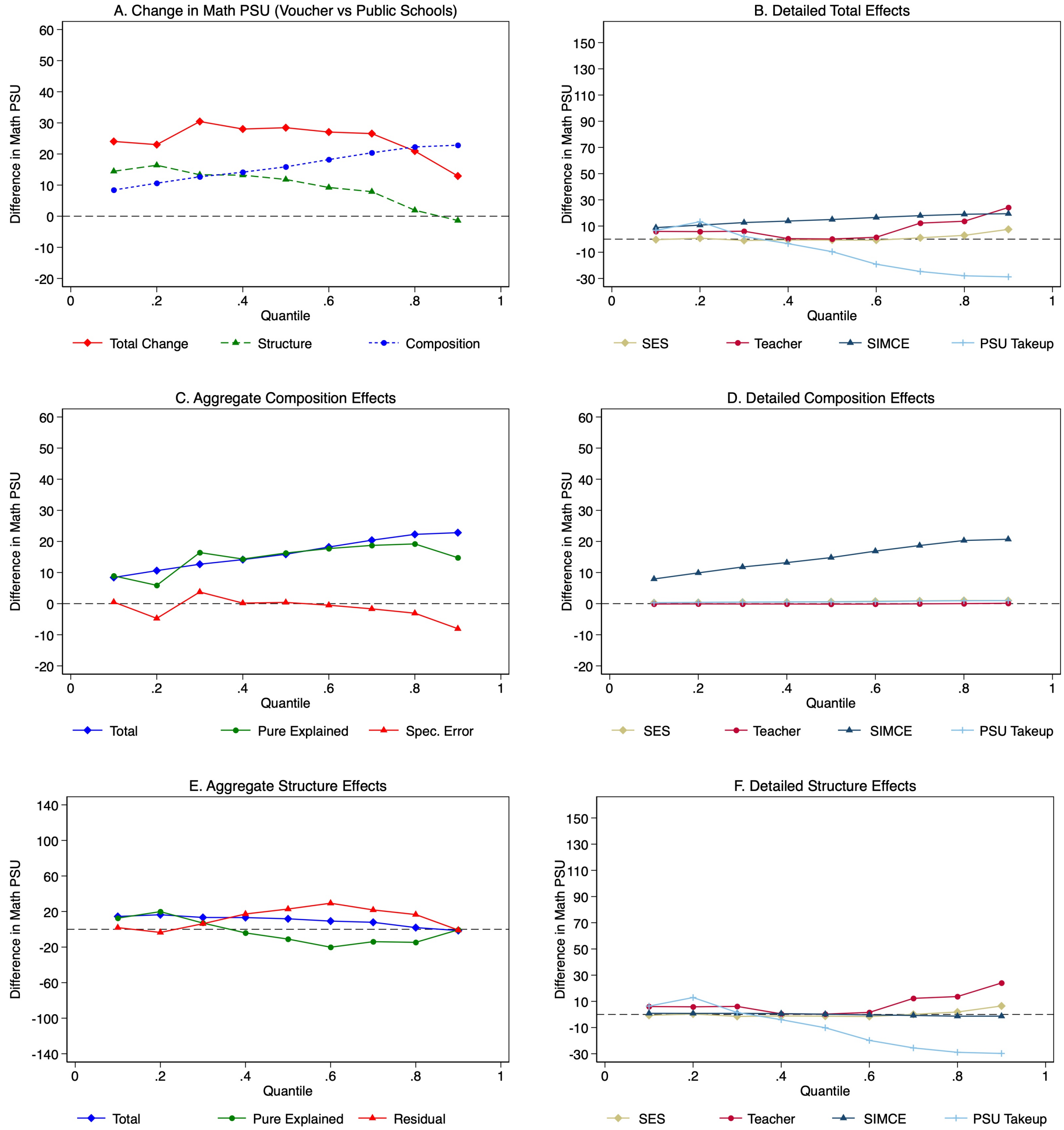
(a) Math PSU



(b) Spanish PSU

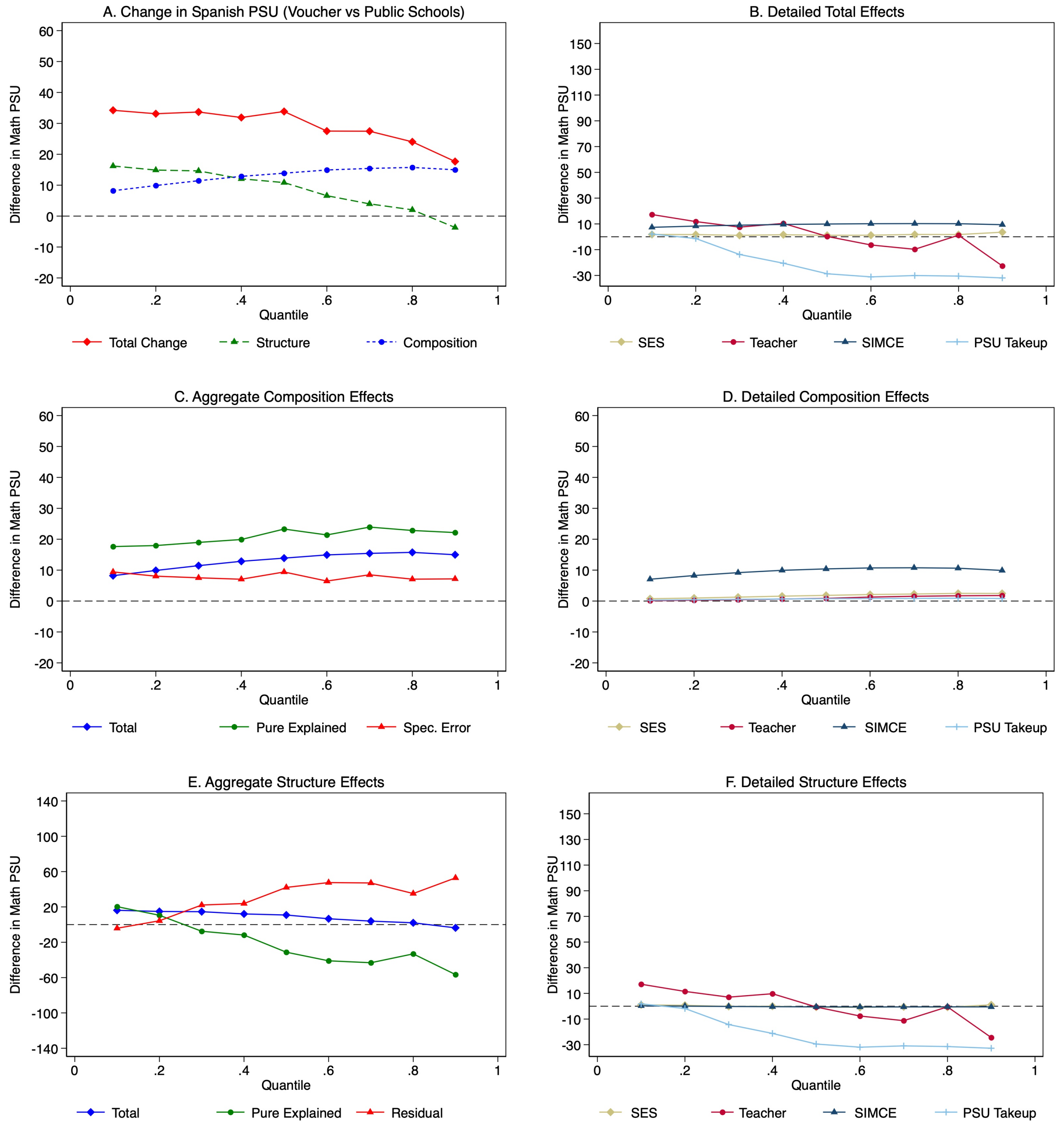
Note: Subfigure a (b) displays the distribution of average teacher math (Spanish) PSU scores computed from our sample teaching 9,938 (9,574) school cohorts of student test takers between the years 2013 and 2021.

Figure 4: Gaps across the Math PSU distribution: Total, Composition, and Structure Effects.



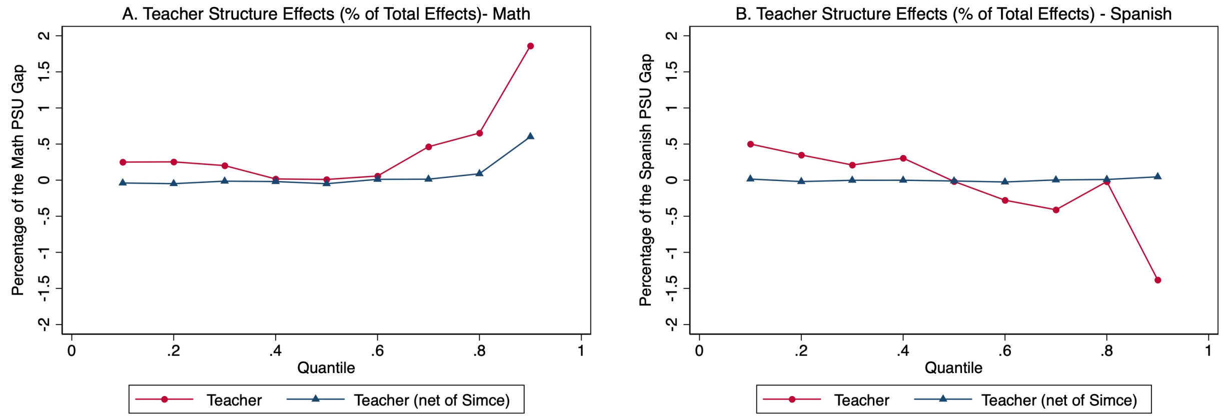
Note: The figure presents the RIF Oaxaca-Blinder decomposition for quantiles of PSU score in math by type of school (public or voucher). Panel A displays the total gap and the portion of it that is explained by the composition and structure effect. Panel B decomposes the total gap by the contribution of each group of variables included in the analysis. Panel C displays the total composition effect and the portion of it that is purely explained and the specification error. Panel D decomposes the purely explained composition effect by the contribution of each group of variables included in the analysis. Panel E displays the total structure effect and the portion of it that is purely explained and residual. Panel F displays decomposes the purely explained structure effect by the contribution of each group of variables included in the analysis. The sample includes 428,973 students covering the period between 2013 and 2021. See Sections 4 and 5 for a formal discussion.

Figure 5: Gaps across the Spanish PSU distribution: Total, Composition and Structure Effects.



Note: The figure presents the RIF Oaxaca-Blinder decomposition for quantiles of PSU score in Spanish by type of school (public or voucher). Panel A displays the total gap and the portion of it that is explained by the composition and structure effect. Panel B decomposes the total gap by the contribution of each group of variables included in the analysis. Panel C displays the total composition effect and the portion of it that is purely explained and specification error. Panel D decomposes the purely explained composition effect by the contribution of each group of variables included in the analysis. Panel E displays the total structure effect and the portion of it that is purely explained and residual. Panel F displays the decomposition of the purely explained structure effect by the contribution of each group of variables included in the analysis. The sample includes 415,315 students covering the period between 2013 and 2021. See Sections 4 and 5 for a formal discussion.

Figure 6: Teachers contribution to structure effect: Total and residualized of previous test score.



Note: The figure presents the contribution of teachers' characteristics to the purely explained structure effect on the PSU score gap and on a residualized measure of the PSU score gap. This residualized measure corresponds to the PSU score minus the effect of the SIMCE test score in the eighth grade. Panel A displays the results for math. Panel B displays the results for Spanish. In both panels, the red line (Teacher) corresponds to the contribution of teachers' variables to the explained structure effect on the PSU gap by quantiles, while the blue line (Teacher res.) corresponds to the contribution of teachers' variables to the explained structure effect on the residualized PSU gap by quantiles. The sample includes 428,973 individuals for math and 415,315 for Spanish, covering the period 2013 to 2021. See Section 5.2 for a formal discussion.