Private School Competition and Student Achievement in Chile

By Alex Tabarrok *

Using Chile's 2005–2024 SIMCE commune panel, this paper asks whether growth in subsidized private-school enrollment raises population achievement or mainly reflects cream skimming. Sorting reshuffles a fixed population and cannot raise its mean, so cream skimming predicts commune mean scores that are flat in private share; in the data they rise—and where they rise is informative. At 4th grade—the grade Hsieh and Urquiola (2006) studied—long differences and panel estimates are small and imprecise, broadly consistent with their null. By 10th grade, after six more years of exposure to the schooling market, the gradient is strong: conditional on a cohort's own 4th-grade achievement, a 10 percentage point increase in private share is associated with a 1.31–1.41 point increase in commune mean scores, positive and significant in every one of ten cohorts spanning two decades of reform—including cohorts schooled after Chile abolished selective admissions—and robust to market definition, sector definition, and cohort-overlap restrictions. The 10th-grade long-difference and value-added estimates imply a per-student benchmark of 0.23–0.28 individual standard deviations, larger in math than in reading. Sorting is present—socioeconomic separation predicts sector score gaps and declining sector means—but it does not produce the aggregate gradient.

* I'd like to thank Miguel Urquiola, Vincent Geloso and Yannick Feussi for comments.

I. Introduction

Does private school competition improve educational outcomes, or does the apparent advantage of private schools simply reflect cream skimming—the selective enrollment of higher-ability students? This question is central to education policy worldwide, and Chile offers a uniquely informative setting to study it.

Chile operates one of the world's most extensive school voucher systems. Since 1981, government-funded vouchers have followed students to both public and private subsidized (*particular subvencionado*, PS) schools, creating a market for education in which PS schools compete directly with public schools for enrollment and funding. Today, over 54% of Chilean students attend PS schools. A further 9.5% attend fully private paid schools (*particular pagado*, PP), which receive no government funding and cater almost exclusively to the wealthiest communes.

A. Relation to the Literature

A large literature beginning with Friedman (1955) argues that vouchers may improve school quality by expanding parental choice and exposing schools to competitive pressure. Subsequent theoretical and empirical work, however, has emphasized that the effects of vouchers are ambiguous because competition may coexist with sorting by income, ability, or preferences over peers (for example, Epple and Romano (1998); Epple, Romano and Urquiola (2017); Hoxby (2003)). MacLeod and Urquiola (2015) show that when schools compete partly on reputation, selective admissions can be privately profitable even when they add no value, formalizing why observed private-school advantages need not reflect productivity. More recent reviews likewise stress both the promise and the heterogeneity of voucher effects. Epple, Romano and Urquiola (2017) and Urquiola (2016) survey the economics literature and conclude that the evidence remains mixed, while Shakeel, Anderson and Wolf (2021), in a meta-analysis of private school voucher studies, report generally positive effects of private schooling on participant outcomes.

The international evidence follows a similar pattern: participant-level effects of voucher programs range from positive—Colombia's PACES lottery (Angrist et al., 2002), Milwaukee (Rouse, 1998), and Washington, DC (Wolf et al., 2013)—to sharply negative in Louisiana (Abdulkadiroğlu, Pathak and Walters, 2018), with Sweden's large-scale reform showing modest positive long-run effects (Böhlmark and Lindahl, 2015). Crucially, participant-level comparisons cannot answer the system-level question, because winners' gains may come at losers' expense through sorting and peer effects. Market-level or aggregate evidence is far scarcer: Muralidharan and Sundararaman (2015) use a two-stage experiment in India to estimate market-level effects of school choice, and Hsieh and Urquiola (2006) study Chilean municipalities. This paper is in that market-level tradition. The tension between participant and population effects is especially central in the Chilean case, where the universal voucher reform created one of the world's largest and longest-running education markets. As a result, Chile has become a key setting for evaluating whether private school expansion raises achievement through productivity and competition, or instead primarily reallocates students across sectors without improving population outcomes.

The Chilean literature has reached mixed conclusions. Early work such as Sapelli (2003) argued that private voucher schools could outperform municipal schools when operating under similar budget constraints, while also stressing the importance of peer effects and sorting in interpreting sectoral differences. The most influential skeptical account is Hsieh and Urquiola (2006), who exploit differential private-sector expansion across municipalities and, importantly, examine mean *population* outcomes rather than only public-private comparisons. Using municipality-level panel data, they find no evidence that generalized school choice improved average test scores, repetition rates, or schooling attainment, and they conclude that the main effect of the reform was increased stratification as higher-performing students moved into the private sector. McEwan, Urquiola and Vegas (2008) similarly characterize the Chilean experience as one in which school choice increased stratification while yielding little clear evidence of gains in average achievement.

Work since Hsieh and Urquiola (2006) has refined both the identification and the picture. Gallego (2013), instrumenting local private entry, finds that inter-school competition raises achievement where entry conditions permit it. Bravo, Mukhopadhyay and Todd (2010) estimate a dynamic structural model and find the 1981 reform raised educational attainment and labor market outcomes. Lara, Mizala and Repetto (2011), using students forced to switch schools between primary and secondary levels and controlling for prior achievement, estimate small private voucher effects relative to earlier cross-sectional estimates. Chumacero, Gómez and Paredes (2011) show that families trade off school quality against distance and price, so demand does respond to quality margins, while Elacqua, Schneider and Buckley (2006) find that stated preferences for quality coexist with revealed sorting on demographics. Mizala and Torche (2012) document the depth of socioeconomic stratification across the three school sectors. A separate strand evaluates the 2008 targeted (SEP) voucher, which raised the subsidy for low-income students and attached school-improvement conditions: Correa, Parro and Reyes (2014) and Murnane et al. (2017) find positive achievement effects, and Neilson (2021) attributes a large share of the gains of poor students to schools competing harder for them once they carried a larger voucher. Taken together, the post-2006 literature is more favorable to competition effects than the original benchmark, but almost all of it studies individual-level or program-specific variation; systemic, population-level evidence over the mature period of the voucher system remains scarce.

B. Contribution and Empirical Strategy

This paper contributes to the debate in two ways. First, it revisits the question with a much longer SIMCE panel spanning 2005-2024, allowing the analysis to cover a broader and more recent period than most prior work. We apply three complementary empirical strategies:

1) **Cross-sectional diagnostic graphs** in the spirit of Tabarrok (2013), which

provide a visual test of the cream skimming hypothesis.

2) **Panel fixed-effects regressions** exploiting within-commune variation in private enrollment shares over two decades.

3) **Cohort value-added models** that control for incoming student quality using lagged 4th-grade scores to assess whether 10th-grade (2M) outcomes remain positively related to private share after prior achievement is taken into account.

The three approaches are presented in increasing order of stringency, and the cohort value-added design is the centerpiece of the paper: by conditioning on the same cohort's own baseline achievement, it estimates the association from baseline-adjusted scores, and because each cohort is a self-contained cross-section, it can be estimated regime by regime across two decades of reform. The cross-sectional diagnostics reject pure cream skimming but also reveal cream skimming signatures, pointing to a mixed picture. The panel and cohort value-added regressions confirm a positive association between PS share and population scores which remains even after controlling further for SES gaps. A consistent theme across designs is a grade gradient: associations are small and imprecise at 4th grade—where Hsieh and Urquiola (2006) looked, and where we broadly reproduce their null—and strong at 10th grade, after cohorts have accumulated six more years of exposure to their schooling market.

The remainder of the paper proceeds as follows. Section II describes Chile's voucher system and the reforms that reshaped it over the sample period, including the mechanics of school selection before and after the 2015 Inclusion Law. Section III describes the data and the construction of the commune-level panel. Section IV develops the cross-sectional diagnostics and compares the theoretical signatures of cream skimming, productivity, and district selection with the data. Section V presents the panel fixed-effects estimates and the SES-gap analysis; Section VI presents the cohort value-added estimates, measures each cohort's exposure to its schooling market grade by grade, and reports cohort-by-cohort stability results. Section VII revisits the Hsieh and Urquiola (2006) benchmark. Section VIII quantifies magnitudes in standard deviation units and weighs the assembled evidence against alternative explanations. Section IX concludes.

II. Institutional Background

A. The 1981 Voucher Reform

Chile introduced a universal flat-rate school voucher in 1981. Under this system, per-student government funding follows the child to any public or subsidized private school (*particular subvencionado*, PS). The reform simultaneously transferred the administration of public schools from the central ministry to municipalities, and it opened the subsidized sector to essentially free entry: historically, very few administrative requirements were needed to open a school and receive the same per-student subsidy as the municipal sector (Bellei and Muñoz, 2023). PS schools are privately operated but publicly funded and may not charge tuition above a regulated copayment, a margin restricted further by the 2015 Inclusion Law. Paid private schools (*particular pagado*, PP) receive no government funding, set their own fees, and serve a different market segment. Public schools are administered by municipalities, municipal corporations, or, since 2018, by Local Public Education Services (SLEPs) that have gradually absorbed municipal systems.

B. Four Decades of Reform

The system our 2005–2024 panel covers is not the static 1981 design. Table 1 summarizes the major reforms; Bellei and Muñoz (2023) provide a full regulatory history. Three changes matter most for interpreting our sample period. First, the 2008 Preferential School Subsidy (SEP) roughly raised the voucher by half for low-income "priority" students and barred participating schools from selecting or charging them, sharply changing the incentive to enroll poor students. Second, the 2015 Inclusion Law prohibited selective admissions and profit-taking in the subsidized sector and began phasing out copayments. Third, the centralized School Admission System (SAE) replaced school-controlled admissions with a lottery-based platform, rolled out region by region between 2016 and 2020. Because these reforms were themselves aimed at the sorting mechanisms that generate cream skimming, the scope for selection differs sharply across our sample period—a feature we exploit below: the panel estimates are shown era by era (Section V) and the value-added estimates cohort by cohort (Section VI) across these policy regimes.

Table 1—Major Reforms to the Chilean Voucher System

Year	Reform
1981	Universal flat per-student voucher; public schools transferred to municipalities; near-free entry for subsidized private providers
1993	Shared financing (*financiamiento compartido*): subsidized private schools may charge regulated copayments
1997	Full School Day reform (JEC): staggered extension of instructional time, tied to infrastructure investment
2008	Preferential School Subsidy (SEP): voucher top-up for low-income students; participating schools may not select or charge them
2011	Quality Assurance System: Agency for Quality of Education (which administers SIMCE) and Superintendency created
2015	Inclusion Law: bans selective admissions and for-profit ownership in the subsidized sector; phases out copayments
2016–2020	Centralized School Admission System (SAE): lottery-based assignment, rolled out region by region
2018–	New Public Education Law: municipal schools gradually transferred to Local Public Education Services (SLEPs)

See Bellei and Muñoz (2023) for a long-run analysis of regulatory models in Chilean education.

C. How Cream Skimming Operated

Before 2015, subsidized private schools could legally and routinely select students. Common instruments included entrance exams and readiness testing for young children, parent interviews, requests for documentation of family background, and—through shared financing—copayments that priced out poorer families. Schools could also counsel out low performers after enrollment. Survey and revealed-preference evidence suggests that while parents cite academic quality as the leading reason for choosing schools, actual choices also track school demographics closely, and demand responds to distance and price (Elacqua, Schneider and

Buckley, 2006; Chumacero, Gómez and Paredes, 2011); the result was deep socioeconomic stratification across sectors (Mizala and Torche, 2012). The Inclusion Law and the SAE dismantled the formal selection instruments: subsidized schools may no longer test, interview, or screen applicants, and oversubscribed schools must accept a centrally run lottery. Copayments are being phased out. Selection on family self-sorting—residential location, information, and willingness to apply—of course remains. This institutional shift matters for our design: if the private-share advantage documented below operated through schools' ability to select students, it should weaken for cohorts schooled after 2016. Section VI shows that it does not.

D. Sector Composition

Table 2 reports the composition of Chilean school enrollment by sector as of 2024.

Table 2—Enrollment by School Sector, 2024

Sector	Share of Enrollment
Municipal	28.9%
SLEP	6.3%
Corp. adm. delegada	1.2%
Total Public	**36.4%**
Particular subvencionado (PS)	54.1%
Particular pagado (PP)	9.5%
Total Private	**63.6%**

Source: Author's calculations from 2024 Ministerio de Educación enrollment records. Public enrollment combines municipal schools, SLEPs, and delegated-administration public schools. Private enrollment combines subsidized private (PS) and fully private paid (PP) schools. Percentages may not sum exactly because of rounding.

III. Data

A. Data Sources

We combine two administrative data sources, both publicly available to researchers:

1) **SIMCE school-level results** (2005–2024): Mean reading and math scores and the number of tested students for each school (identified by its national school code, *rbd*), by grade and year. SIMCE is Chile's national standardized assessment, administered near-universally to the tested grades. The school-level public databases are distributed by the Agencia de Calidad de la Educación through its online information center; we use every available public wave for grades 4B (4th grade) and 2M (10th grade).

2) **Mineduc enrollment files** (2004–2024): Grade-specific enrollment by school, with dependency codes identifying the school's sector (public, PS, or PP). These summary enrollment files are published by the Ministerio de Educación's Centro de Estudios on its open-data portal.

The two sources link at the school level through the *rbd* code. A replication package containing all ingestion, harmonization, and estimation code accompanies the paper; year-specific file formats and the commune-code harmonization are documented there.

SIMCE is anchored to a scale with mean 250 and individual standard deviation 50 in the base year for each test series, and scores are equated across years using item response theory so that they are comparable over time within a grade-subject series. The scale is *not* re-standardized each year: national means can and do drift, so the level of the population mean is informative rather than fixed by construction. That said, none of our results depend on intertemporal comparability of the national level. All panel specifications include grade $\times$ subject $\times$ year fixed effects, so estimates are identified from cross-commune differences within a test-year and are unaffected by common additive shifts in the national level of the scale (a year-specific multiplicative rescaling would change the units of within-year cross-commune differences, but not their sign or ranking); the cross-sectional diagnostics in Section IV likewise compare communes to one another. Throughout the paper we report effect sizes both in SIMCE points and in standard deviation units, using the individual test SD of 50 and the commune-level SD of mean scores of about 18 (Table 4).

B. *Analysis Sample and Sector Definitions*

The private share varies dramatically across communes, from 0% (rural communes with no private schools) to over 94% (affluent urban communes like Las Condes). Table 3 shows the school-level SES classification (*cod_grupo*) by school type: public, private subsidized (PS), or fully private (PP). Public schools are concentrated in the bottom SES categories but there is common support with PS schools in all but the top SES category, where public enrollment is essentially absent. In contrast, PP schools are almost entirely concentrated in the top category. In the pooled 2005–2024 4B and 2M data, 95.8% of PP enrollment is in the highest SES group, compared with 1.4% for PS schools and essentially zero for public schools.

Table 3—SES Composition by School Sector

Sector	Low (*Bajo*)	Lower middle (*Medio bajo*)	Middle (*Medio*)	Upper middle (*Medio alto*)	High (*Alto*)
Public	26.0%	48.2%	22.7%	3.1%	0.0%
PS	7.8%	26.0%	40.1%	24.7%	1.4%
PP	0.0%	0.0%	0.0%	4.2%	95.8%

Source: Author's calculations using SIMCE school socioeconomic classifications (*cod_grupo*) matched to Ministerio de Educación enrollment data, pooling grades 4B and 2M over 2005–2024. Public refers to public schools, PS to subsidized private schools, and PP to fully private paid schools. Percentages are weighted by tested students within sector and may not sum exactly because of rounding.

Because PP schools operate in a fundamentally different market—charging fees, receiving no government funding, drawing from the highest SES groups, and geographically concentrated in a handful of wealthy communes—we exclude them from the analysis. For the same reasons, we exclude 10 rich communes where the PP share is over 50 percent.[1] Our focus is on competition between public and private schools within the Chilean voucher system. The exclusions are conservative: including PP schools and the high-PP communes *raises* the estimates (Appendix Table C1). We exclude them to put cream skimming on the strongest possible footing and to keep the market definition clean.

The PS-public score gap is persistent across the sample period. Pooling all years and grades, PS schools average roughly 260 in reading and 260 in math, compared to 244 and 237 for public schools—gaps of approximately 16 and 23 points respectively.

C. Construction of Commune-Level Panel

School-level SIMCE scores are aggregated to commune means using tested-student counts as weights:

$$\bar{Y}_{cgst} = \frac{\sum_{i \in c} Y_{igst} \cdot N_{igst}}{\sum_{i \in c} N_{igst}} \tag{1}$$

where i indexes schools in commune c, grade g, subject s, year t; reading and math means are aggregated separately, each using its own subject-specific tested-student counts. The PS enrollment share is:

$$\text{share\_priv}_{cgt} = \frac{\text{enroll}^{PS}_{cgt}}{\text{enroll}^{Public}_{cgt} + \text{enroll}^{PS}_{cgt}} \tag{2}$$

[1]Six are affluent Greater Santiago communes (Huechuraba, La Reina, Las Condes, Lo Barnechea, Providencia, Vitacura) and a seventh, Calera de Tango, is on the rural fringe of the Metropolitan Region. The remaining three are Machalí, adjacent to Rancagua, and two small Valparaíso-region communes (Panquehue and Calle Larga) in which a single paid school accounts for the majority of enrollment in at least one grade-year cell.

The resulting panel contains 19,221 commune × grade × subject × year observations spanning 17 SIMCE years (2005–2024; the 2019 wave is unavailable for these grades because testing was disrupted by the October 2019 social protests, and the 2020–2021 waves were cancelled during the COVID-19 pandemic), 336 communes, and two grade levels (4B and 2M).[2]

Table 4 reports summary statistics for the resulting commune-level panel.

TABLE 4— SUMMARY STATISTICS: COMMUNE-LEVEL PANEL

	Mean	SD	Min	Max	N
Commune mean SIMCE score	247.49	17.68	148.00	376.00	19221
Public-school mean SIMCE score	240.05	18.40	142.00	314.00	13760
Private-school mean SIMCE score	258.05	19.58	155.44	350.00	13760
Private-public score gap	18.00	23.11	-73.00	143.00	13760
Public enrollment share (%)	66.00	27.75	0.00	100.00	19221
Private enrollment share (%)	34.00	27.75	0.00	100.00	19221
Commune SES mean (1-5)	2.05	0.63	1.00	5.00	19221
Low-SES share (%)	71.12	28.66	0.00	100.00	19221
Public-school SES mean (1-5)	1.77	0.52	1.00	4.00	13760
Private-school SES mean (1-5)	2.55	0.73	1.00	5.00	13760
Private-public SES gap	0.77	0.64	-1.60	3.00	13760

Source: Author's calculations using SIMCE commune-level panel data. Paid private (PP) schools are excluded. Sector-specific rows are based on commune-level cells with at least 10 tested students in both public and private-subsidized categories, so N varies by row.

IV. Cross-Sectional Evidence: Exploratory Diagnostics

A. The Population-Mean Diagnostic

In the spirit of exploratory data analysis, Tabarrok (2013) proposes a visual diagnostic for distinguishing cream skimming from non-compositional effects. Under a pure cream skimming model, private schools selectively enroll higher-quality students but add no value. Thus, instead of looking at private and public students and trying to control for student quality, Tabarrok (2013) looks at the mean population

[2]Commune codes are harmonized to the current standard across the 2007 regional reform, the 2010 creation of Marga Marga province, and the 2018 creation of the Ñuble region, so communes whose codes changed over the period are tracked as single units.

score as the private share increases across districts. If the cream skimming model holds then, all else equal, the mean population score cannot vary with the private share because the population mean is not influenced by sorting Hsieh and Urquiola (2006).

Two features of this diagnostic deserve emphasis. First, the market-level population mean is the right outcome for the system-level question because, as Hsieh and Urquiola (2006) argue, it registers the net effect of school choice on every student in the market: gains to students who move into private schools are counted together with whatever choice does to the students who remain behind, whether through the loss of strong peers or through competitive pressure on public schools. Effects that merely redistribute achievement across a fixed population of students cancel in the population mean; what survives is what school choice adds on net—private productivity advantages, competitive responses, and any peer effects that are not zero-sum. Sector-level comparisons cannot make this distinction: a private-public score gap may reflect productivity or merely selective enrollment. Second, the flat population line under cream skimming is not an assumed counterfactual—it is the *testable prediction* of the sorting-only null. Sorting reshuffles a fixed population of students across sectors, and no reshuffling, however selective, can change the population mean. This prediction holds regardless of the specific selection mechanism (exams, interviews, prices, or self-selection), which is what makes the diagnostic robust to not observing how selection operates.

B. Signature Patterns: Cream Skimming, Productivity and District Selection

For a simple model of cream skimming, assume that student quality is drawn from a uniform distribution on $[0, 1]$ and that private schools over-draw above the z-th percentile by a factor β(see the appendix for details). Figure 1 illustrates the signature pattern of cream skimming. The private mean exceeds the public mean because private schools draw disproportionately from the top of the public-school distribution. As the private share of schooling increases, the private-school mean must decline because there is less cream left to skim. The public-school mean must also decline with rising private share because a larger share of high-scoring students has already sorted into the private sector. Notice, however, that the population mean is exactly flat at 250: cream skimming or sorting of any kind cannot, by itself, raise or lower the population mean. We will look for each of these signatures in the data.

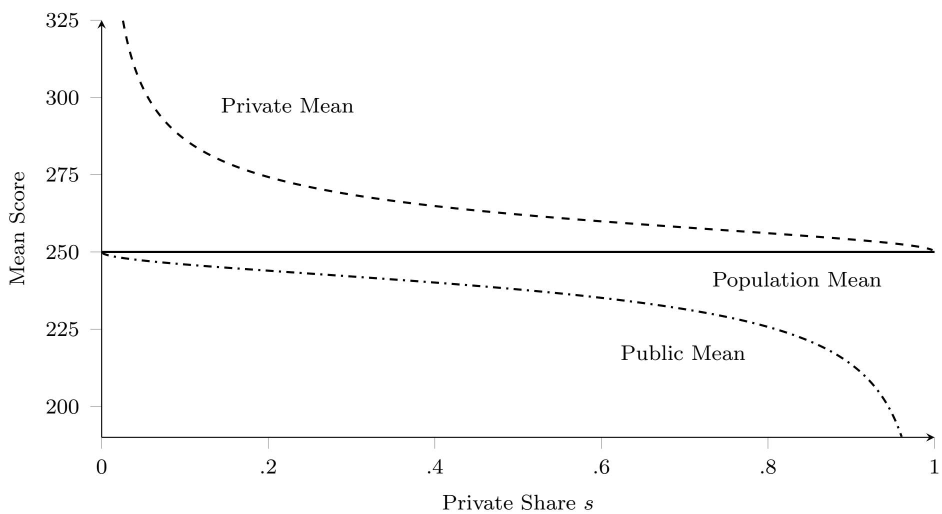

Figure 1. The Signature Cream-Skimming Pattern: In the model, the population mean score remains constant as private share rises because private schools have no productivity effect. The private-school mean declines with private share because, as enrollment expands, there is less cream left to skim. The public-school mean also declines because a larger share of high-scoring students has already sorted into the private sector. The model is loosely calibrated to a population mean of 250 and a 25-point private-public gap at private share $s = 0.62$, roughly matching Chile.

In contrast, consider the signature pattern of the productivity model shown in Figure 2. In this model, students are sorted randomly but private schools are more productive, that is, they raise scores directly. The key pattern is that the private and public means are constant with respect to private share, while the population mean increases as more students enter the higher-productivity private sector.

Finally, consider a third model of interest: district selection. In this case, the

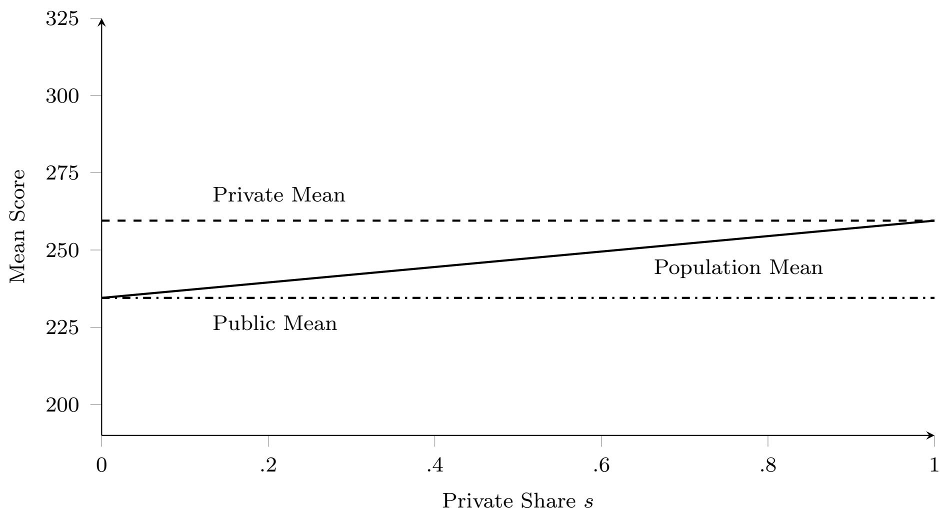

FIGURE 2. THE SIGNATURE PRODUCTIVITY PATTERN: THE PRIVATE-SCHOOL MEAN EXCEEDS THE PUBLIC-SCHOOL MEAN BY A CONSTANT AMOUNT BECAUSE PRIVATE SCHOOLS RAISE ACHIEVEMENT DIRECTLY RATHER THAN BY SKIMMING HIGHER-PERFORMING STUDENTS. AS PRIVATE SHARE RISES, SECTOR MEANS REMAIN CONSTANT WHILE THE POPULATION MEAN RISES BECAUSE MORE STUDENTS ENTER THE HIGHER PRODUCTIVITY SECTOR. THE MODEL IS LOOSELY CALIBRATED TO A POPULATION MEAN OF 250 AND A 25-POINT PRIVATE-PUBLIC GAP AT PRIVATE SHARE $s = 0.62$, ROUGHLY MATCHING CHILE.

districts in which the private share of education is large also have higher income and higher-achieving students overall. Figure 3 illustrates the signature pattern. In this model, the population mean increases with private share, but so do the private and public means. Observationally, that pattern can be hard to distinguish from a case in which competition raises productivity in both sectors.

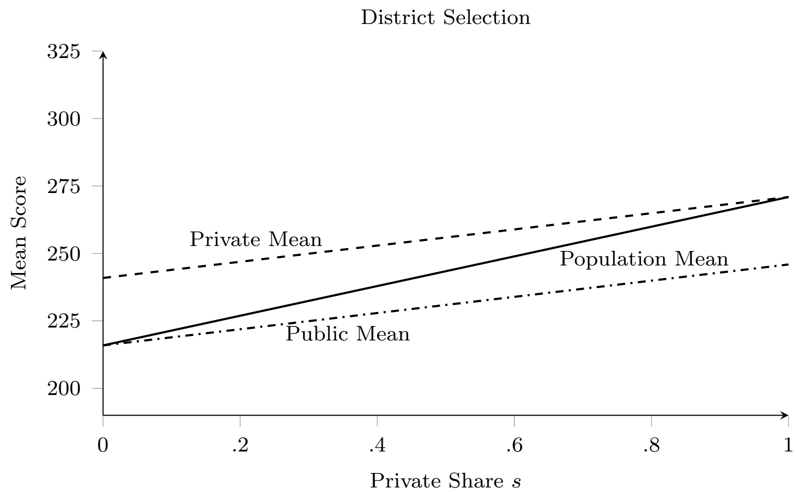

Figure 3. The Signature District Selection Pattern: In this model, the population mean rises with the private share, but so do the private and public means, in all cases driven by correlation between the private share and student quality by district. The model is loosely calibrated to a population mean of 250 and a 25-point private-public gap at private share $s = 0.62$, roughly matching Chile.

C. Data Patterns

Figures 4 and 5 plot smoothed commune-level SIMCE means against PS enrollment share, pooling all years separately by subject. The two subjects show distinct patterns. In reading, the population mean rises modestly over part of the range but is broadly flat and eventually declines, while both sector means fall as PS share increases. Overall, reading looks closer to the cream-skimming pattern. In math, by contrast, the population mean rises clearly across the range of PS shares, from roughly 242 to 257, while the sector means flatten or decline. Math therefore looks closer to the productivity model, though the declining sector means indicate that sorting remains present.

These figures are descriptive: they pool two decades of data and mix cross-commune differences with changes over time. We therefore use them only to illustrate the contrasting signatures in the data; the panel and cohort designs that follow provide the paper's evidence.

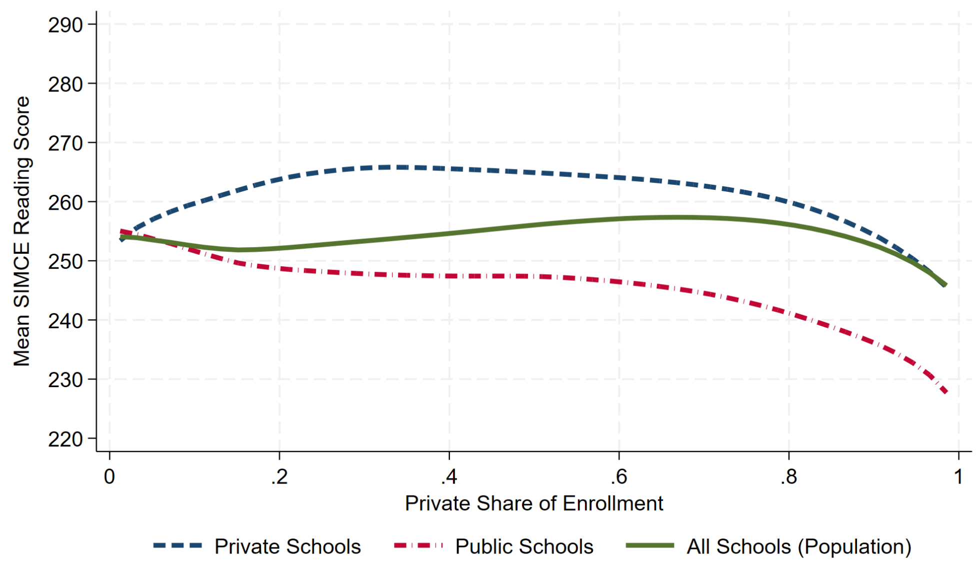

FIGURE 4. READING: COMMUNE-LEVEL SIMCE MEANS BY PRIVATE ENROLLMENT SHARE, ALL YEARS POOLED (2005–2024). LOCAL POLYNOMIAL FITS WEIGHTED BY TESTED STUDENTS.

V. Panel Fixed-Effects Regressions

A. Specification and Identifying Variation

To further test for cream skimming, we turn to a panel fixed-effects specification:

$$Y_{cgst} = \alpha_c + \gamma_{gst} + \beta \cdot \text{share\_priv}_{cgt} + \varepsilon_{cgst} \tag{3}$$

where Y_{cgst} is the tested-student-weighted mean SIMCE score across all public and PS students in commune c, grade g, subject s, and year t—the population

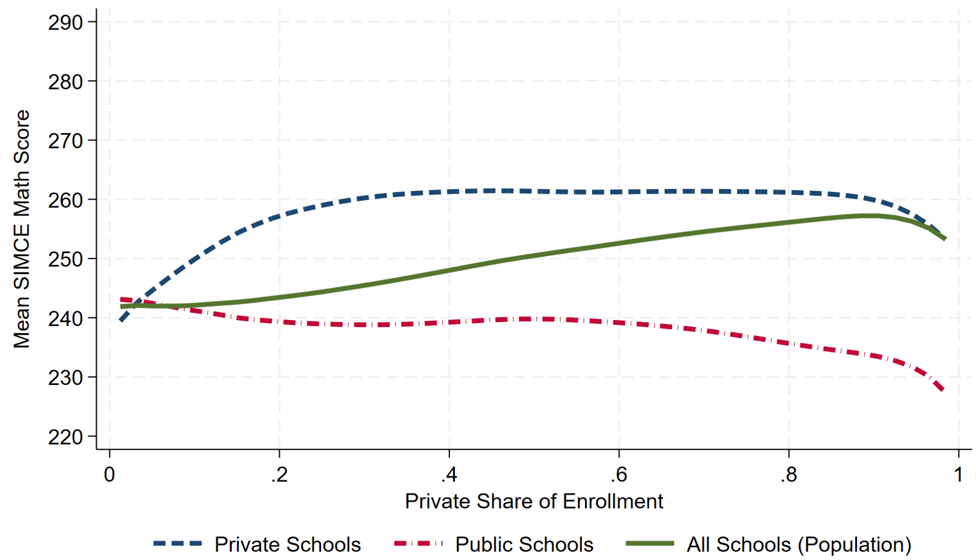

FIGURE 5. MATH: COMMUNE-LEVEL SIMCE MEANS BY PRIVATE ENROLLMENT SHARE, ALL YEARS POOLED (2005–2024). LOCAL POLYNOMIAL FITS WEIGHTED BY TESTED STUDENTS.

mean of equation (1)—and share_priv_{cgt} is the PS share of enrollment defined in equation (2). α_c is a commune fixed effect (absorbing time-invariant commune characteristics) and γ_{gst} is a grade $\times$ subject $\times$ year fixed effect (absorbing nationwide shocks to each test). All regressions are weighted by the number of tested students and cluster standard errors at the commune level.

It is useful to examine the temporal variation directly. In the balanced 4B sample, changes between successive observed SIMCE waves are typically small, averaging about 0.56 percentage points, but cumulative change over the 2005–2024 long difference is substantial: commune-level changes range from about -30 to $+54$ percentage points, with a median increase of about 7 percentage points, as shown in Figure 6.

This period coincided with sub-replacement fertility in Chile, partly offset by immigration, as well as important changes in the organization of the school market. A natural concern is that the estimated within-commune variation mainly reflects shrinking or shifting student populations rather than reallocation across sectors.

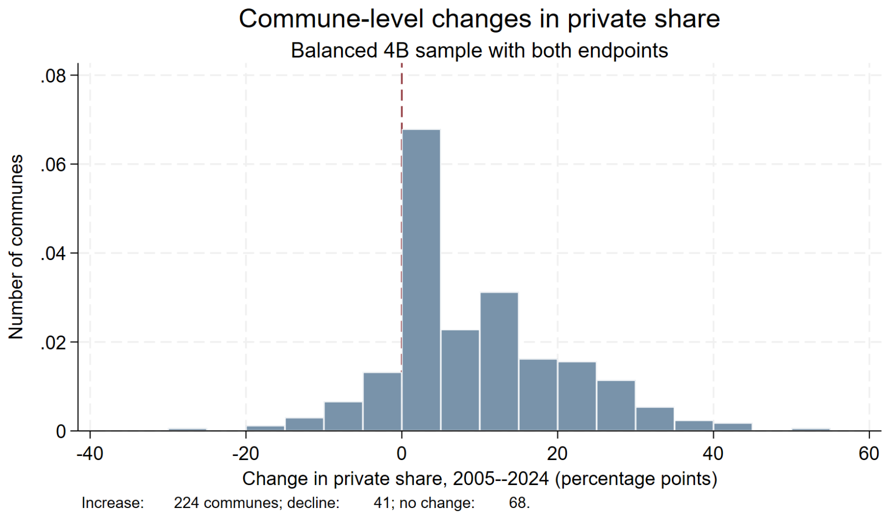

Figure 6. Within-commune changes in private share over the mature period. Changes between successive observed SIMCE waves are usually small, but cumulative commune-level reallocation from 2005 to 2024 is substantial.

Descriptively, however, changes in private share are only weakly correlated with total enrollment change (0.096) and with baseline low-SES share (-0.140), but are much more strongly correlated with public enrollment decline (-0.465) and private enrollment growth (0.505). These correlations are partly mechanical—the share is a ratio of sector enrollments—but they show the within-commune variation reflects sectoral reallocation rather than overall enrollment decline.[3]

B. Results

Table 5 reports the panel estimates. Column 1, the baseline, gives $\hat{\beta} = 10.5$ (SE = 2.9): within commune, a 10 percentage point increase in PS share is associated with about a one-point increase in the population mean score.

Columns 2–3 split the sample by grade and reveal a gradient that recurs throughout the paper: the association is small and imprecise at 4th grade (4.3, SE = 4.1) and stronger at 10th grade (6.2, SE = 3.0), where cohorts have spent six more years

[3]Appendix B reports additional descriptive evidence on the panel variation.

in their schooling market. Columns 4–5 split by subject: math (11.6, SE = 3.7) exceeds reading (9.4, SE = 2.3), so the subject asymmetry documented below in the cohort design is present here as well. Column 6, discussed in Section V.C, replaces commune with commune × grade fixed effects.

The panel estimates show a positive relationship between PS share and population scores after accounting for fixed commune characteristics and common grade × subject × year shocks. The relationship is concentrated at grade 10 and is larger in math than in reading.

TABLE 5—PANEL FIXED EFFECTS: BASELINE, GRADES, SUBJECTS, AND ROBUSTNESS

	(1) Commune FE	(2) 4B only	(3) 2M only	(4) Reading	(5) Math	(6) C×G FE
share_priv	10.52***	4.32	6.17**	9.41***	11.63***	5.42**
	(2.90)	(4.10)	(3.01)	(2.29)	(3.71)	(2.74)
N	19221	11293	7927	9609	9610	19220
Clusters	336	336	315	335	335	336
R^2	0.737	0.840	0.761	0.804	0.707	0.813

Sample: Full commune panel after excluding PP schools and communes with PP share at or above the threshold in the full school market.
Unit of observation: commune × grade × subject × year. All columns include grade × subject × year fixed effects (grade × year in the single-subject columns).
Weights: number of tested students. Standard errors are clustered by commune.
Column 1 is the baseline specification with commune fixed effects. Columns 2–3 split the sample by grade and columns 4–5 by subject, all with commune fixed effects. Column 6 replaces commune with commune × grade fixed effects, so the slope is identified only from changes within a commune-grade market over time (Section V.C).

C. Robustness

We assess the robustness of the panel relationship along four dimensions: the fixed-effects structure, the geographic definition of school markets, stability across policy regimes, and whether measured socioeconomic sorting accounts for the private-share slope.

Fixed Effects

Commune fixed effects absorb persistent differences across communes in demographic composition and student background, including average student preparedness, parental income, and parental education. Because 4th- and 10th-grade students are drawn from the same commune population, these characteristics should be broadly common to both grades. The baseline specification therefore uses a single fixed effect for each commune.

Commune $\times$ grade fixed effects relax this assumption by allowing every commune to have a separate intercept for grades 4 and 10. This controls for persistent differences between a commune's primary- and secondary-school markets, including grade-specific school supply, commuting to secondary-school hubs, and selection at the primary-to-secondary transition. The cost is that it removes all persistent cross-grade variation within communes, leaving only changes within a commune-grade market over time. Discarding this source of variation removes roughly half of the identifying variation and magnifies attenuation from measurement error in private shares. In column 6 of Table 5, the estimate falls from 10.5 to 5.4 (SE = 2.7) but remains positive and significant. The commune $\times$ grade estimate also lies between the separate 4th- and 10th-grade estimates of 4.3 and 6.2, while the baseline estimate is larger because it additionally uses the cross-grade margin.

Market Definition

To account for commuting at the primary-to-secondary transition, we also broaden the geographic definition of the school market. Most students attend school in the commune where they live (Hsieh and Urquiola, 2006), but Greater Santiago is an important exception: its dense conurbation permits substantial movement across commune boundaries. Appendix Table C2 therefore re-estimates the models after treating Greater Santiago as a single market and, separately, after excluding the Metropolitan Region.

The baseline commune fixed-effects estimate remains strong under both definitions: 8.4 (SE = 2.8) with Greater Santiago aggregated and 10.3 (SE = 3.1) with the Metropolitan Region excluded, compared with 10.5 in the main specification. Thus, the baseline result is not driven by treating Santiago's interconnected communes as separate school markets. The commune × grade estimates are smaller and imprecise under the same variations (2.3 and 1.8), indicating that the short-run temporal association in that specification is concentrated in Greater Santiago (Appendix C).

Within a single grade, commune and commune × grade fixed effects yield the same estimator. The grade-specific estimates in columns 2–3—and therefore the grade gradient central to the paper—do not depend on the fixed-effects choice.

Stability Across Policy Eras

The panel spans four distinct policy regimes: the pre-SEP baseline (2005–2009), the SEP era (2010–2014), the Inclusion Law era (2015–2018), and the post-COVID years under fully national centralized admissions (2022–2024). If any single reform—or a confounder tied to one—drove the association, the slope should break at the corresponding date. Table 6 estimates the baseline specification separately within each era, each regression with its own commune fixed effects, so every slope is identified only from variation inside that era. The separate-regression estimates (column 1) are 6.2, 14.0, 14.1, and 9.2; all positive, the last three significant, the pre-SEP baseline imprecise (a short window with the least within-market share movement). A pooled specification with era-specific slopes and common commune fixed effects (column 2)—which forces every era to share one fixed effect and therefore leans on longer-run variation—gives slopes of 9.1–12.4, all significant at the 1 percent level, and exact equality of the four slopes is not rejected ($p = 0.11$). The essential point holds in both parameterizations: every era slope is positive, and there is no break at any reform date.

Beyond stability, the ordering is informative: schools' formal ability to select students was greatest before 2015 and abolished afterward, yet the slope is, if anything, larger in the Inclusion Law era than in the pre-SEP baseline—the opposite of what a selection-driven association would predict. The panel evidence thus matches the cohort-by-cohort evidence of Section VI: the association appears in every regime of the mature voucher system, not as the artifact of a particular reform.

Table 6—Panel Estimates by Policy Era

Era	Separate regressions	Pooled, common FE	N (separate)
2005-2009	6.15 (4.58)	9.09*** (3.20)	4,494
2010-2014	13.96** (5.47)	10.53*** (3.20)	5,736
2015-2018	14.09*** (4.24)	12.43*** (2.98)	5,122
2022-2024	9.17*** (2.98)	9.80*** (2.75)	3,869

Coefficients on PS share by era, weighted by tested students, standard errors (in parentheses) clustered by commune. Column 1: separate regressions per era, each with its own commune and grade × subject × year fixed effects. Column 2: one pooled regression with era-specific slopes and common commune fixed effects, which leans on longer-run variation; a joint test of equal slopes across eras gives $p = 0.110$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

SES Sorting and the Private-Share Slope

To investigate the extent of cream skimming, we note that communes differ in the variance of within-commune SES: some communes are uniformly poor, while others span a wide socioeconomic range. We therefore construct commune-level measures of sector-specific SES from the school-level `cod_grupo` classification (a five-point socioeconomic index assigned to each school). For each commune × grade × year cell, we compute the tested-student-weighted mean SES separately for public and private schools, then define the *SES gap* $\Delta S \equiv \bar{S}_{\text{priv}} - \bar{S}_{\text{pub}}$. A larger ΔS indicates greater socioeconomic separation between sectors.

Communes with a greater SES gap have more room for cream skimming, although ΔS is an equilibrium object rather than a pure measure of schools' selection opportunities—a large gap can also reflect stronger demand for private schooling among better-off families. We begin by asking how the private-public score gap varies with the SES gap. Figure 7 provides direct evidence on this point: the private-public score gap in each commune-year is plotted against the corresponding SES gap. While there is a clear positive relationship—communes with larger SES gaps between sectors also have larger score gaps—the wide dispersion around the linear fit shows that measured SES separation is far from a sufficient statistic for the score gap: many communes show substantial gaps even at small measured SES differences. Appendix Table E1 formalizes the figure: regressing the private-public score gap (ΔY) on ΔS within commune, weighting each cell by the harmonic mean of the two sectors' tested-student counts (which down-weights cells where either sector mean is imprecisely estimated), a one-unit increase in ΔS is associated with an 18.2-point increase in the score gap ($p < 0.01$). Socioeconomic separation is a strong predictor of how far apart the two sectors' scores sit—which is what makes it a natural moderator for the test that follows.

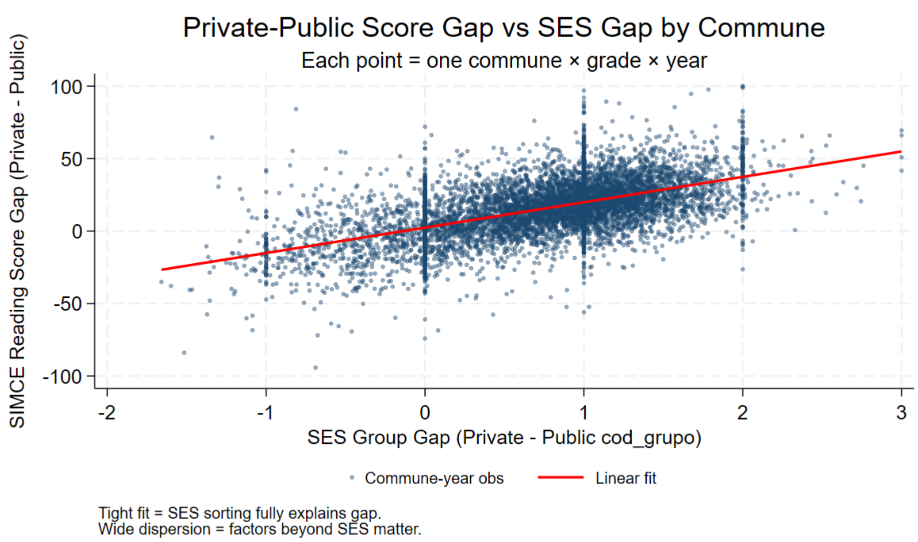

FIGURE 7. PRIVATE-PUBLIC SCORE GAP VS. SES GAP BY COMMUNE-YEAR, READING. EACH POINT IS ONE COMMUNE × GRADE × YEAR OBSERVATION. IF MEASURED SES SORTING FULLY EXPLAINED THE SCORE GAP, POINTS WOULD CLUSTER TIGHTLY ALONG THE FITTED LINE; THE WIDE DISPERSION SHOWS IT IS FAR FROM A SUFFICIENT STATISTIC. THE MATH VERSION IS SIMILAR (APPENDIX FIGURE E4).

If the positive population-score slope documented above were an artifact of sorting, it should be strongest where sorting has the most room to operate. To test this, we estimate

$$(4)\quad Y_{cgst} = \alpha_c + \gamma_{gst} + \beta \cdot \text{share\_priv}_{cgt} + \rho \cdot \Delta S^c_{cgt} + \theta \cdot \left(\text{share\_priv}_{cgt} \times \Delta S^c_{cgt}\right) + \varepsilon_{cgst}$$

where ΔS^c is the SES gap centered on its sample mean, so that β is the private-share slope at the average level of SES separation and θ measures how that slope varies with separation. Constructing sector-specific SES requires both sectors to be observed with sufficient support in the same commune $\times$ grade $\times$ year cell, so these regressions are estimated on a common-support subsample rather than the full panel of Table 5.

Because ΔS^c is centered, β is the slope at the average level of separation, so the sharp predictions concern how the slope varies with ΔS. If the population slope reflected sorting-related confounding, it should rise with the scope for sorting ($\theta > 0$) and fall toward zero in communes whose sectors are not socioeconomically separated at all (a slope near zero at $\Delta S = 0$). If instead the slope reflects productivity or competition, it should not depend on the scope for selection: $\theta \approx 0$, with a positive slope even at $\Delta S = 0$.

Table 7 reports the population-score results. The private-share slope is stable across specifications: $\hat{\beta} = 8.5$ with no SES controls (column 1), 8.4 with ΔS as a control (column 2), and 8.4 at the mean ΔS in the interaction specification (column 3), with $\hat{\theta} = 1.0$ (SE = 3.5)—no detectable dependence of the slope on the scope for sorting. The implied slope at $\Delta S = 0$, in communes with no socioeconomic separation between sectors, is 7.6 (SE = 4.3, $p = 0.08$): close to the slope at the mean gap and positive, though less precisely estimated at this edge of the data. Replacing the scalar ΔS with the full vector of categorical SES-composition gaps and their interactions leaves the private-share coefficient essentially unchanged (Appendix Table E3), so the result is not an artifact of the particular summary measure of separation.[4] These results match the predictions of productivity or competition, not those of sorting: measured SES separation does not explain the slope. Re-estimating the interaction within the cohort value-added design gives the same answer (Appendix Table E4); sorting the five-category index misses is taken up by the cohort design of Section VI.

Table 7—Population Scores, Private Share, and SES Gap Interaction

	(1) Baseline	(2) + ΔS	(3) Interaction
Private Share	8.515***	8.382***	8.438***
	(3.054)	(2.907)	(2.876)
SES Gap		0.165	
		(0.817)	
SES Gap (Centered)			-0.328
			(1.951)
Share × SES Gap			0.962
			(3.495)
Commune FE	Yes	Yes	Yes
GxSxT FE	Yes	Yes	Yes
N	13760	13760	13760
Clusters	261	261	261
R2	0.755	0.755	0.755

Sample: Common-support commune panel with both public and PS sectors observed well enough to construct sector-specific SES measures.
Unit of observation: commune × grade × subject × year.
Weights: number of tested students.
Fixed effects and clusters are reported in the table.
Standard errors are clustered by commune.

Sorting is far from absent, however—it simply shows up where the accounting says it must: within sectors rather than in the population mean. Appendix Table E2 re-estimates equation (4) with the sector means as dependent variables, weighting by each sector's own tested students. Within commune, a rising PS share is associated with a falling public-sector mean (-12.3, $p < 0.05$) and a flat-to-falling private-

[4]These regressions use commune fixed effects, matching the baseline specification. With commune × grade fixed effects, the common-support baseline is 5.1 (SE = 3.1), and 5.2 (SE = 3.1) with the ΔS control—consistent with column 6 of Table 5, though less precise on this restricted sample.

sector mean (-3.9, n.s.), and both declines are substantially steeper in communes with large SES gaps (interaction coefficients of -19.0 and -10.3, both $p < 0.01$). This is precisely the compositional signature visible at high private shares in the diagnostics of Section IV: expansion drains stronger students from the public sector and dilutes the private sector, most visibly where sorting is strong. Yet these within-sector declines coexist with a population slope that is positive and flat in ΔS: as an accounting matter, enrollment shifts toward the sector whose mean is roughly 18 points higher, which outweighs the within-sector declines. Sorting reshuffles and dilutes within sectors; it does not produce the aggregate slope.

D. Distributional Effects: The Bottom Tail

Mean effects can hide who gains. SIMCE classifies each tested student against absolute proficiency standards; we translate the official categories (*Insuficiente*, *Elemental*, *Adecuado*) as below standard, elementary, and at standard. The commune-level share of students scoring below standard measures the size of the low-achievement tail directly. School-level proficiency distributions are available from 2014, so we re-estimate the panel design on a 2014–2024 window with the proficiency shares as outcomes.

Table 8 reports the results. Within commune, a 10 percentage point increase in PS share is associated with a 0.9 percentage point decline in the share of students below standard ($\hat{\beta} = -9.4$, SE $= 2.8$, on a base of 43.5 percent), with the now-familiar subject asymmetry: -10.7 in math versus -8.1 in reading. The share at standard rises by about half as much ($+4.7$, SE $= 1.7$), so the gains are concentrated in moving students out of the failing category rather than in enlarging the top. With commune $\times$ grade fixed effects the estimate remains negative but is imprecise (-2.3, SE $= 3.2$, column 2): the proficiency data begin only in 2014, and a one-decade window contains little within-market share variation, so the commune-FE estimates should be read as primarily descriptive.

These results also sharpen the interpretation of the mean effects. Like the population mean, the population-wide share below an *absolute* standard is invariant to reshuffling: selection cannot reduce the number of low-performing students simply by reallocating a fixed tested population across schools. Moreover, the below-standard share falls by twice as much as the at-standard share rises, indicating that gains extend to weak students rather than accruing only at the top. The shrinking bottom tail therefore points to improved school inputs reaching the students most at risk.

TABLE 8—DISTRIBUTIONAL OUTCOMES: PROFICIENCY SHARES AND PRIVATE SHARE, 2014–2024

	(1) % Below std.	(2) C×G FE	(3) Reading	(4) Math	(5) % At std.
Private share	-9.36***	-2.29	-8.05***	-10.65***	4.70***
	(2.76)	(3.17)	(2.35)	(3.42)	(1.74)
Observations	10,106	10,106	5,053	5,051	10,106
Communes	334	334	333	333	334

Standard errors in parentheses
Below standard = *Insuficiente*; at standard = *Adecuado*.
Columns 1–4: below-standard share (3–4 by subject); column 5: at-standard share.
Column 2: commune × grade fixed effects; other columns: commune fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Taken together, the panel results establish a grade gradient that is stable across fixed-effects structures, market definitions, policy regimes, and SES specifications, with complementary evidence that the below-standard share falls as private enrollment rises. The cohort design next conditions directly on each cohort's own 4th-grade achievement when explaining its 10th-grade outcome.

VI. Cohort Value-Added Evidence

A. Motivation

The panel fixed-effects approach controls for time-invariant commune characteristics. An alternative and arguably more informative strategy exploits the fact that students take SIMCE at two points in their schooling: 4th grade (4B) and 10th grade (2M), six years apart. The 4B commune mean serves as a control for incoming student quality.

This design is powerful for three reasons. First, it controls for selection up to the baseline directly rather than by assumption: whatever combination of family background, ability, motivation, and prior sorting determined a cohort's achievement by 4th grade is embodied in its 4B score, so the private-share coefficient is identified from achievement conditional on that baseline rather than from levels. Unobserved determinants of baseline achievement—the usual Achilles' heel of school-choice comparisons—are absorbed to the extent they operate through 4th-grade performance; selection occurring after 4th grade is not, and we return to it below. Second, the six-year window ties the coefficient to exposure: the 10th-grade share proxies for the schooling environment the cohort passed through between the two tests, and Section VI.E replaces the proxy with the directly measured grade-by-grade path. Third, because each cohort is a self-contained cross-section conditional on its own baseline, the design needs no pooling across policy regimes to be identified—which is what allows the cohort-by-cohort estimates reported below.

B. Cohort Construction

We match 4B scores from year t with 2M scores from year $t+6$—the same student cohort observed at two points in its schooling. We label each cohort by its pair of test years: "4B 2006 → 2M 2012," for example, denotes the cohort tested in 4th grade in 2006 and in 10th grade in 2012. Ten cohorts can be matched, the earliest being 4B 2006 → 2M 2012 and the latest 4B 2018 → 2M 2024. The sequence has a three-cohort gap in the middle: because no SIMCE was administered in 2019–2021, the cohorts that would have reached 2M in those years (4B years 2013–2015) cannot be matched. The 4B 2005 cohort is also unmatched, for a different reason: no 2M test was given in 2011, as the 2M assessment was biennial before 2012. The merged dataset contains 6,142 commune × subject × cohort observations in the baseline value-added sample.

C. *Specification*

The value-added regression is:

$$Y^{2M}_{cs,t+6} = \beta_1 \cdot \text{share\_priv}^{2M}_{c,t+6} + \beta_2 \cdot Y^{4B}_{cs,t} + \gamma_{st} + \varepsilon_{cs,t} \tag{5}$$

where s indexes the subject, $Y^{4B}_{cs,t}$ is the same-subject 4B commune mean, γ_{st} is a subject $\times$ cohort fixed effect (absorbing national trends and test difficulty shifts), and t indexes the cohort's 4B year, so the same cohort is observed again at 2M in year $t+6$.

D. *Results*

Table 9 reports the cohort value-added results. The main result (column 1) shows that after controlling for the same cohort's 4B scores, a 10 percentage point increase in the 2M PS share is associated with a **1.4 SIMCE point** improvement in 2M scores ($\hat{\beta} = 14.05$, SE $= 2.47$). The 4B coefficient is 0.73, confirming that 4B scores are a strong proxy for incoming student quality.

Column 2 adds the commune's PS share when the cohort was in 4th grade; it adds nothing conditional on the 10th-grade share (-1.47, SE $= 4.79$), whose coefficient is unchanged ($14.05 \to 14.94$). This rules out the narrower story that only the primary-era market matters. Section VI.E takes the question further by measuring the cohort's exposure directly.

Columns 3–4 split by subject; we defer their discussion to the heterogeneity analysis below.

E. *Cumulative Exposure*

The 10th-grade share in Table 9 is a proxy for the treatment of interest: the schooling market the cohort experienced between its two tests. Because the enrollment files report grade-specific enrollment in every year, including 2019–2021, we

TABLE 9—COHORT VALUE-ADDED: 2M SCORE ON PRIVATE SHARE, CONTROLLING FOR 4B BASELINE

	(1) VA	(2) VA+4Bshr	(3) Reading	(4) Math
share_priv_2m	14.05***	14.94***	9.114***	18.40***
	(2.471)	(3.833)	(1.948)	(3.156)
score_4b	0.734***	0.740***	0.555***	0.841***
	(0.0708)	(0.0634)	(0.0590)	(0.0827)
share_priv_4b		-1.472		
		(4.785)		
Commune FE	No	No	No	No
Other FE	Subject x cohort	Subject x cohort	Cohort	Cohort
Observations	6,142	6,142	3,070	3,072
Clusters	315	315	315	314
R2	0.513	0.513	0.405	0.495

Sample: Matched 4B-to-2M cohorts after excluding PP schools and high-PP communes in the full school market.
Unit of observation: commune × subject × cohort.
Weights: number of tested students in 2M.
Fixed effects and clusters are reported in the table. Standard errors are clustered by commune.

can measure that exposure path directly. Let $\text{share\_priv}_{c,g,y}$ denote the PS share of enrollment in commune c's grade-g market in year y, the grade-level analogue of equation (2). A cohort taking 4B in year t occupies grade g in year $t + (g - 4)$, so its cumulative exposure is

$$\text{exposure}_{ct} = \frac{1}{6} \sum_{g=5}^{10} \text{share\_priv}_{c,\,g,\,t+(g-4)}. \tag{6}$$

Measuring treatment as cumulative exposure increases the coefficient from 14.1 for the 10th-grade endpoint to 15.2 for the full grades 5–10 exposure average. Table 10 compares alternative exposure windows on a common sample. Beginning with the grade-4 share, which precedes the exposure window, the estimate rises monotonically from 12.4 as successive grades are added, reaching 14.6 through grade 9 and 15.2 through grade 10. Reparameterizing exposure as the baseline level and the within-window change yields statistically indistinguishable coefficients (13.5 and 14.9): a percentage point of private share has a similar association with achievement whether the cohort begins with it or acquires it along the way, as a duration-of-exposure model predicts.

The timing comparison also reduces concern about reverse causality: endpoint enrollment could respond to same-year test outcomes. The grade-4 share and the grades 5–9 exposure average are predetermined. The latter yields 14.6 (SE = 3.3), slightly above the endpoint estimate, so the result does not depend on measuring the market in the test year.

These comparisons have two limitations. The timing measures are highly correlated—the correlation between the grade-4 share and the full exposure average is 0.94—so differences among their coefficients are suggestive rather than sharp. Moreover, measuring cumulative exposure does not address migration or endogenous school placement; those concerns are examined in the cohort-fidelity analysis below and the market-definition robustness checks in Appendix C.

Table 10—Cohort Exposure Paths: Alternative Treatment Timings

	(1) Grade 4	(2) Exposure	(3) Level + change	(4) Pre-outcome	(5) Grade 10
PS share, grade 4	12.41***		13.46***		
	(3.36)		(3.11)		
Exposure, grades 5–10		15.17***			
		(3.18)			
Change, grades 4–10			14.94***		
			(3.83)		
Exposure, grades 5–9				14.57***	
				(3.29)	
PS share, grade 10					14.05***
					(2.47)
Baseline score (4B)	0.70***	0.71***	0.74***	0.71***	0.73***
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
Observations	6,142	6,142	6,142	6,142	6,142
Communes	315	315	315	315	315

Standard errors in parentheses
All columns control for the same-subject 4B commune mean and subject × cohort fixed effects, weighted by 2M tested students.
Pre-outcome: average share over grades 5–9, measured entirely before the outcome year.
Standard errors clustered by commune. Common sample.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

F. Robustness

We assess the cohort design's sensitivity to the baseline adjustment, imperfect cohort matching, commune fixed effects, and socioeconomic sorting.

Sensitivity to the baseline adjustment.

The design leans on the 4B control, so we verify that nothing depends on how that control enters. Appendix Table D3 replaces the pooled linear control with progressively more flexible functions of the baseline—subject-by-cohort-specific linear slopes, quadratics, restricted cubic splines, and within-cell quintile indicators—and the private-share estimate ranges from 11.9 to 14.0, significant at the 1 percent level throughout. Even constraining the persistence coefficient to unity, a gain-score specification, gives 13.4. Persistence is also already permitted to vary across subjects and cohorts elsewhere in the analysis: the subject-specific columns estimate separate 4B coefficients, and the cohort-by-cohort regressions below estimate one for every cohort, with the private-share coefficient positive and significant throughout.

Cohort fidelity.

The match is a cohort-level pseudo-panel, not a literal student panel: grade repetition shifts students into later cohorts, some students drop out or miss the test, students move across communes—most importantly at the primary-to-secondary transition, when students in small communes often travel to secondary schools elsewhere—and, late in the sample, immigration adds students who were not tested at 4B. The tested-count ratio $n_{2M}(t+6)/n_{4B}(t)$ summarizes the net effect. Nationally the ratio is 0.82–0.91 for most cohorts; the median commune cell retains about 80 percent of its 4B count. The tails reveal the geography of the secondary transition: 16 percent of cells have ratios below one-half (small communes exporting students to secondary schools elsewhere) and 3 percent above 1.5 (the receiving hubs), but these low-fidelity cells are small—cells with ratios between 0.5 and 1.5

contain 87 percent of all tested 2M students. Imperfect overlap makes the 4B mean a noisy proxy for the 2M cohort's true baseline, which attenuates the 4B coefficient and could in principle contaminate the private-share estimate. Appendix Table D2 shows it does not: restricting the sample to cells with ratios in $[0.5, 1.5]$ raises the private-share estimate from 14.05 to 15.47, and a strict $[0.75, 1.25]$ band raises it to 16.79. The estimate is not driven by the low-overlap cells; if anything, they pull it down.

Commune fixed effects.

Adding commune fixed effects to equation (5) yields 0.40 (SE = 4.39). This is uninformative rather than contradictory: within commune, across the ten matched cohorts, the standard deviation of the 2M private share is only 4.1 percentage points (against 22.7 raw), so that specification discards nearly all identifying variation. The within-commune evidence at horizons long enough to contain real share variation is the panel of Section V and the long differences of Table 11.

SES-gap adjustment.

Re-estimating the SES-gap specification of Section V within the cohort design also leaves the main result unchanged. On the common-support sample, adding the interaction leaves the slope at the mean SES gap essentially unchanged (13.08 to 13.04), while the interaction is positive but imprecisely estimated (6.19; Appendix Table E4). Socioeconomic sorting therefore does not account for the private-share association.

G. Heterogeneity Across Cohorts and Subjects

Cohort-by-cohort estimates.

Because the value-added design is identified within each cohort's cross-section, we can also estimate the PS-share coefficient separately for each of the ten matched cohorts. Figure 8 plots the estimates against the 2M outcome year, with the pooled estimate as a dashed reference line; Appendix Table D1 reports the underlying numbers. Every cohort-specific estimate is positive and statistically significant, ranging from 10.8 to 18.0 points. A joint test rejects exact equality of the slopes ($p = 0.002$), so the magnitude does vary across cohorts, but the variation is unsystematic: there is no trend and no break at the major policy events. In particular, the three post-COVID cohorts (2M years 2022–2024), who entered secondary school under centralized admissions (SAE) and were schooled through the pandemic, show estimates of 11.2–15.3, well inside the range of the pre-2018 cohorts. The stability of the association across the SEP, Inclusion Law, SAE, and COVID regimes weighs against explanations tied to any single reform, though SAE and pandemic effects cannot be separately identified at the 2018–2022 gap. The subject-specific series (Appendix Figure D1) shows the math coefficient exceeding the reading coefficient in every pre-COVID cohort, with the two converging in the post-COVID cohorts.

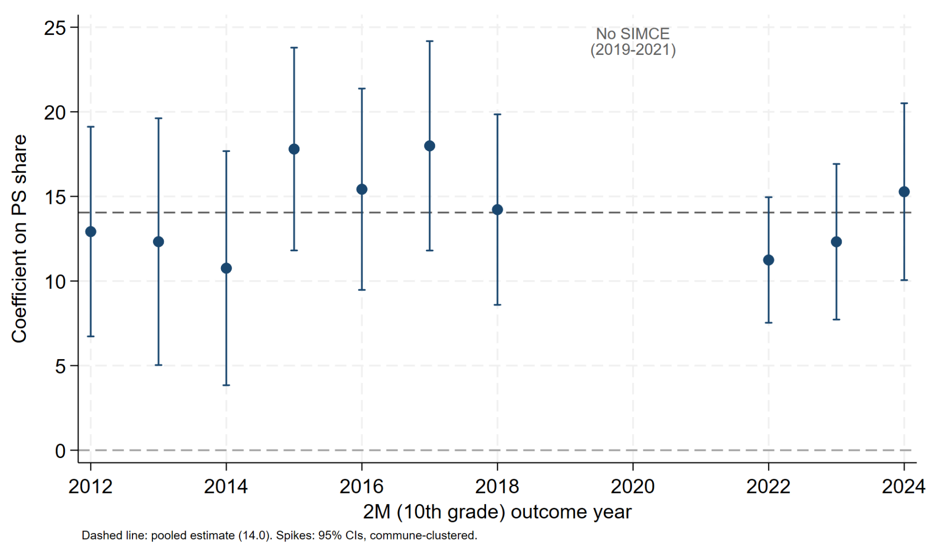

Figure 8. Cohort-specific value-added estimates. Each point is a separate regression of commune 2M scores on the 2M PS share and the same cohort's 4B scores, pooling subjects with subject fixed effects and weighting by 2M tested students. Spikes are 95% confidence intervals based on commune-clustered standard errors. The dashed line is the pooled estimate (14.05). No SIMCE was administered in 2019–2021.

Subject heterogeneity.

Columns 3–4 of Table 9 split by subject. The prior-achievement-adjusted association is roughly twice as large for math (18.40) as for reading (9.11). The direction of this asymmetry is what the education production literature would predict: mathematics is learned primarily at school, while reading skill draws heavily on the home environment—parental vocabulary, books, and everyday language exposure—and is therefore less sensitive to differences in school inputs. Consistent with this, estimates of teacher and school effects on test scores are systematically larger in math than in reading (Chetty, Friedman and Rockoff, 2014; Kraft, 2020). A math-to-reading ratio of about two is thus a coherence check on the interpretation of the coefficient as a schooling effect rather than a spurious correlation, which would have no reason to respect the subject asymmetry.

VII. Revisiting the Hsieh-Urquiola Benchmark

Hsieh and Urquiola (2006) also measure outcome using population means but they study the broad expansion era after the 1981 reform, when private entry and sorting were both accelerating. Our data begin in 2005, when the market is already mature and we cover a longer time and bigger cross-sectional sample. The original HU regressions cover only 94 communes.

Hsieh and Urquiola (2006) use long differences rather than a pure panel setup, and—a detail that matters for the comparison—their test-score outcomes are entirely 4th-grade scores: the PER of 1982 and the SIMCE of 1988 and 1996, which in their period tested 4th graders in even years and 8th graders in odd years. The closest replication of their design on the mature era is therefore a 4th-grade long difference. Column 1 of Table 11 estimates it: the change in a commune's 4B mean from 2005 to 2024 on the change in its 4B-market private share, controlling for the commune's baseline score and baseline private share, as they did. Column 2 applies the identical design at 10th grade (2006–2024). Column 3 reports the panel estimate on the 10th-grade subsample (the 2M column of Table 5), and column 4 the cohort value-added estimate with lagged 4th-grade scores.

Table 11—Revisiting Hsieh-Urquiola on the Later SIMCE Panel

	(1) LD 4B (H–U margin)	(2) LD 2M	(3) Panel 2M	(4) Cohort VA
share_priv	4.16	11.45**	6.17**	14.05***
	(4.46)	(5.67)	(3.01)	(2.47)
Baseline controls	Yes	Yes	No	Yes
Lag 4B score	No	No	No	Yes
N	333	294	7927	6142
Clusters	333	294	315	315
R2	0.437	0.474	0.761	0.513

Columns 1–2 are grade-specific commune long differences (4B: 2005–2024; 2M: 2006–2024), each controlling for the commune's baseline score and baseline private share, weighted by average tested students. Column 3 is the panel specification of Section V estimated on the 10th-grade subsample, with commune and subject × year fixed effects. Column 4 is the cohort value-added specification of Section VI.
Standard errors are clustered by commune.

The comparison refines rather than overturns the benchmark. At 4th grade, their null broadly persists in the mature era: the long difference is 4.16 (SE = 4.46)—positive but small and imprecise, much as they found for the expansion era. The departure is at 10th grade, which their test data never measured: the 2M long difference is 11.45 (SE = 5.67), the 10th-grade panel estimate is 6.17 (SE = 3.01), and the cohort value-added estimate—which conditions on the same cohort's 4th-grade achievement—is 14.05 (SE = 2.47). The mature Chilean equilibrium thus departs from the H-U null precisely where cohorts have accumulated the most exposure to the voucher market, and looks most like it where they measured: at a grade whose students have spent only a few years in school.

VIII. Discussion

The results raise two broader questions: how large are the estimated effects, and how well does the joint evidence distinguish a productivity effect from the leading alternatives?

A. Magnitudes

Across the main specifications, the estimated effect of PS share on the commune-wide mean SIMCE score is organized by grade and design rather than clustered around a single number: 4.2 (imprecise) in the 4th-grade long difference and 4.3 in the 4th-grade panel; at 10th grade, 6.2 in the panel, 11.5 in the long difference, and 14.1 (13.1 on the SES-restricted sample) in the cohort value-added design. Because effect sizes in the education literature are conventionally reported in standard deviation units, Table 12 expresses each estimate three ways. On the individual test scale (SD = 50 by construction), a 10 percentage point increase in PS share raises the commune population mean by 0.008–0.009 test SD at 4th grade, and at 10th grade by 0.012 test SD in the panel and 0.023–0.028 test SD in the long-difference and value-added designs. Relative to the distribution of commune mean scores (SD = 17.7; Table 4), the 10th-grade changes are 0.04–0.08 commune-level SD; commune means are far less dispersed than individual scores, so effect sizes measured against them are mechanically larger.

The third conversion turns the population-mean effect into a per-student benchmark. A 10 percentage point increase in PS share moves one student in ten across sectors, so if the gains accrue to the students who move—that is, if competition and peer effects on non-movers are small—the implied effect of attending a PS school is ten times the per-10-point population effect. If some of the gain instead comes through competition on non-movers, the per-student effect is smaller; the conversion is an upper bound under those assumptions. Read in these units, the gradient in Table 12 maps onto exposure: about 0.08–0.09 SD (imprecise) at 4th grade, and at 10th grade 0.12 SD in the panel and 0.23–0.28 SD in the long-difference and value-added designs. Moving a commune from the 25th to the 75th percentile of PS share—roughly 30 percentage points—implies a commune-mean gain of about 1.9 points in the 10th-grade panel and 3.4–4.2 points in the long-difference and value-added designs.

Table 12—Effect Sizes in Standard Deviation Units

			Per 10 pp of PS share		Per student
Specification	Coef.	SE	Test SD	Commune SD	Test SD
Long difference, grade 4 (H–U's grade)	4.16	(4.46)	0.008	0.024	0.08
Panel, grade 4	4.32	(4.10)	0.009	0.024	0.09
Long difference, grade 10	11.45	(5.67)	0.023	0.065	0.23
Panel, grade 10	6.17	(3.01)	0.012	0.035	0.12
Cohort value added (grade 10 outcome)	14.05	(2.47)	0.028	0.079	0.28
SES-restricted cohort value added	13.08	(2.92)	0.026	0.074	0.26

Coefficients are the PS-share estimates from the main text, ordered by grade and design: the grade-specific long differences and panel estimates (Tables 11 and 5), then the cohort value-added designs (Tables 9 and E4). Within a single grade the commune and commune × grade fixed-effects estimators coincide, so the panel rows do not depend on that choice. Test SD is the individual SIMCE standard deviation (50 by test construction). Commune SD is the standard deviation of commune mean scores across commune-grade-subject-year cells in the panel estimation sample (17.7; see Table 4). The per-10-pp columns give the change in the commune population mean for a 10 percentage point increase in PS share. The per-student column divides the coefficient by the individual SD: it is the implied effect on a student who moves from the public to the PS sector, under the assumption that competition and peer effects on non-movers are small.

The per-student benchmark is the number most comparable to the literature. Under the assumption of small competition effects, the change in the population mean with the private share is best interpreted as a productivity effect from moving a student from a lower-productivity public school to a higher-productivity private school—at 10th grade, 0.23–0.28 of an individual-level SD in the long-difference and value-added designs (final column of Table 12). By Kraft (2020)'s benchmarks, a per-student effect of that size falls right at the boundary between "medium" and "large"—placing it in the top tier of education interventions. Chetty, Friedman and Rockoff (2014) estimate that a 1 SD increase in teacher value added in a single grade raises earnings by about 1 percent at age 28. A 1 SD increase in teacher VA roughly corresponds to 0.1 SD in student test scores so raising student achievement by 0.20 SD results in a 2 percent increase in annual lifetime earnings on average (Kraft, 2020). As another comparison, the Tennessee STAR class-size experiment—reducing classes from 22 to 15 students—produced effects of roughly 0.2 SD in math and reading but at considerable expense (Krueger, 1999). A voucher system has a similar effect at low marginal fiscal cost since voucher funding is on par with public school funding.

B. *Weighing the Evidence Against the Alternatives*

The affirmative case rests on the joint pattern of results rather than any single estimate. The association is small and imprecise at 4th grade but large at 10th grade (Table 12), as cumulative exposure would predict. It is consistently larger in math than in reading, does not systematically vary with the socioeconomic separation between sectors, and appears across policy regimes and cohorts. Taken together, the grade gradient, subject asymmetry, SES results, and stability over time are difficult to reconcile with simple selection or a common shock.

Sorting alone cannot move the mean of a fixed population of tested students. The empirical population, however, consists of tested public and subsidized-private students attending schools in the commune, so migration, cross-commune attendance, movement into fully private schooling, grade progression, or differential test participation could change the observed mean. The evidence makes such changes unlikely to explain the central result. Changes in private share are weakly related to total enrollment change and more closely associated with reallocation between sectors (Appendix B). The cohort estimate strengthens when the retained Greater Santiago communes are aggregated into one market or the Metropolitan Region is excluded, and when the sample is restricted to cells with stable tested cohort sizes. It also remains strong when excluded communes and fully private schools are included (Appendix C). These checks cannot eliminate demographic change, but they make population composition alone an unlikely explanation for the repeated cohort relationship.

The remaining concern is an omitted determinant of achievement correlated with private-school exposure. To confound the cohort estimate, it must arise—or begin to affect achievement differentially—during the six years between 4th and 10th grade, vary across communes with subsequent private-school exposure, and not already be reflected in baseline achievement. Subject-by-cohort fixed effects absorb shocks common to Chile in each period, and the relationship appears across ten overlapping six-year windows. A viable confounder must therefore persist or recur across communes and cohorts while reproducing the grade gradient, subject asymmetry, and stability across selection regimes. No instrument is available, so a slowly changing local factor cannot be formally excluded. The parsimonious reading is nevertheless that greater exposure to the subsidized private sector raised achievement in the mature Chilean voucher system.

IX. Conclusion

The question of whether private-school competition improves educational outcomes or merely reshuffles students has been central to education policy since Friedman (1955). Much of the literature attempts to separate productivity from selection by controlling for observed student characteristics, but that approach is difficult when important determinants of selection are unobserved. Following Hsieh and Urquiola (2006) in Chile and Tabarrok (2013) in India, this paper instead examines how mean population scores vary with private-school share. With a fixed population, pure cream skimming cannot by itself raise the population mean, so this approach does not require knowing student characteristics or their influence on scores and provides a useful diagnostic when private enrollment varies widely across markets.

Hsieh and Urquiola (2006) found no increase in mean population outcomes during the early expansion of Chile's voucher system in the 1980s and early 1990s; their test-score evidence is for 4th graders. Using a longer and larger panel (2005–2024), we find a pattern in the later equilibrium that refines rather than overturns theirs. At 4th grade—the grade they studied—our estimates remain, like theirs, small and imprecise. At 10th grade, after cohorts have accumulated six more years of exposure to their schooling market, the association is strong: 11.5 in a long difference and 14.1 in the cohort value-added design that conditions on the same cohort's 4th-grade achievement, positive and significant in every one of ten cohorts across two decades of reform. A 10 percentage point increase in PS share raises the commune-wide 10th-grade mean by about 1.1–1.4 points; under the assumption of small competition effects on non-movers, this corresponds to a 0.23–0.28 SD increase in individual scores from attending the private sector—a meaningful gain, on par with smaller class sizes from the Tennessee Project STAR experiment.

Sorting remains an important part of Chile's voucher market: sector means decline at high PS shares, and socioeconomic separation predicts sector score gaps. But these compositional effects do not account for the rising population mean at grade 10. The evidence therefore points to a mature system in which selection and productivity coexist: private expansion reshuffles students across sectors, while greater cumulative exposure to the subsidized-private market is associated with higher achievement, especially in math and among students at the bottom of the distribution.

Declaration of AI Use

During the preparation of this work, I used ChatGPT Pro and Claude to assist with data and replication scripts and to review the manuscript for errors and weaknesses. I reviewed and edited all outputs and take full responsibility for the content of the article.

REFERENCES

Abdulkadiroğlu, Atila, Parag A. Pathak, and Christopher R. Walters. 2018. "Free to Choose: Can School Choice Reduce Student Achievement?" *American Economic Journal: Applied Economics*, 10(1): 175–206.

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review*, 92(5): 1535–1558.

Bellei, Cristóbal, and Gonzalo Muñoz. 2023. "Models of Regulation, Education Policies, and Changes in the Education System: A Long-Term Analysis of the Chilean Case." *Journal of Educational Change*, 24(1): 49–76.

Böhlmark, Anders, and Mikael Lindahl. 2015. "Independent Schools and Long-Run Educational Outcomes: Evidence from Sweden's Large-Scale Voucher Reform." *Economica*, 82(327): 508–551.

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.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633–2679.

Chumacero, Rómulo A., Daniel Gómez, and Ricardo D. Paredes. 2011. "I Would Walk 500 Miles (If It Paid): Vouchers and School Choice in Chile." *Economics of Education Review*, 30(5): 1103–1114.

Correa, Juan A., Francisco Parro, and Loreto Reyes. 2014. "The Effects of Vouchers on School Results: Evidence from Chile's Targeted Voucher Program." *Journal of Human Capital*, 8(4): 351–398.

Elacqua, Gregory, Mark Schneider, and Jack Buckley. 2006. "School Choice in Chile: Is It Class or the Classroom?" *Journal of Policy Analysis and Management*, 25(3): 577–601.

Epple, Dennis, and Richard E. Romano. 1998. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review*, 88(1): 33–62.

Epple, Dennis, Richard E. Romano, and Miguel Urquiola. 2017. "School Vouchers: A Survey of the Economics Literature." *Journal of Economic Literature*, 55(2): 441–492.

Friedman, Milton. 1955. "The Role of Government in Education." In *Economics and the Public Interest.* , ed. Robert A. Solo, 123–144. New Brunswick, NJ:Rutgers University Press.

Gallego, Francisco A. 2013. "When Does Inter-School Competition Matter? Evidence from the Chilean "Voucher" System." *The B.E. Journal of Economic Analysis & Policy*, 13(2).

Hoxby, Caroline M. 2003. "School Choice and School Productivity: Could School Choice Be a Tide that Lifts All Boats?" In *The Economics of School Choice.* , ed. Caroline M. Hoxby, 287–342. Chicago:University of Chicago Press.

Hsieh, Chang-Tai, and Miguel Urquiola. 2006. "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program." *Journal of Public Economics*, 90(8–9): 1477–1503.

Kraft, Matthew A. 2020. "Interpreting Effect Sizes of Education Interventions." *Educational Researcher*, 49(4): 241–253.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions*." *The Quarterly Journal of Economics*, 114(2): 497–532.

Lara, Bernardo, Alejandra Mizala, and Andrea Repetto. 2011. "The Effectiveness of Private Voucher Education: Evidence from Structural School Switches." *Educational Evaluation and Policy Analysis*, 33(2): 119–137.

MacLeod, W. Bentley, and Miguel Urquiola. 2015. "Reputation and School Competition." *American Economic Review*, 105(11): 3471–3488.

McEwan, Patrick J., Miguel Urquiola, and Emiliana Vegas. 2008. "School Choice, Stratification, and Information on School Performance: Lessons from Chile." *Economía*, 8(2): 1–27.

Mizala, Alejandra, and Florencia Torche. 2012. "Bringing the Schools Back In: The Stratification of Educational Achievement in the Chilean Voucher System." *International Journal of Educational Development*, 32(1): 132–144.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. "The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India." *The Quarterly Journal of Economics*, 130(3): 1011–1066.

Murnane, Richard J., Marcus R. Waldman, John B. Willett, Maria Soledad Bos, and Emiliana Vegas. 2017. "The Consequences of Educational Voucher Reform in Chile." National Bureau of Economic Research NBER Working Paper 23550.

Neilson, Christopher A. 2021. "Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students." Princeton University Working Paper.

Rouse, Cecilia Elena. 1998. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *The Quarterly Journal of Economics*, 113(2): 553–602.

Sapelli, Claudio. 2003. "The Chilean Voucher System: Some New Results and Research Challenges." *Cuadernos de Economía*, 40(121): 530–538.

Shakeel, M. Danish, Kaitlin P. Anderson, and Patrick J. Wolf. 2021. "The Participant Effects of Private School Vouchers around the Globe: A Meta-Analytic and Systematic Review." *School Effectiveness and School Improvement*, 32(4): 509–542.

Tabarrok, Alexander. 2013. "Private Education in India: A Novel Test of Cream Skimming." *Contemporary Economic Policy*, 31(1): 1–12. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1465-7287.2011.00286.x.

Urquiola, Miguel. 2016. "Competition Among Schools: Traditional Public and Private Schools." In *Handbook of the Economics of Education*. Vol. 5, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 209–237. Amsterdam:Elsevier.

Wolf, Patrick J., Brian Kisida, Babette Gutmann, Michael Puma, Nada Eissa, and Lou Rizzo. 2013. "School Vouchers and Student Outcomes: Experimental Evidence from Washington, DC." *Journal of Policy Analysis and Management*, 32(2): 246–270.

A Simple Cream-Skimming Model

This appendix presents a simple model showing how cream skimming alone can generate a private-school advantage in average test scores even when private schools have no causal effect on achievement.

Let student ability or test performance be denoted by X, where

$$X \sim U[0, 1].$$

Thus the population mean is

$$E[X] = \frac{1}{2}.$$

Suppose the probability that a student attends a private school depends on ability:

$$\Pr(\text{private} \mid X = x) = a + \beta \mathbf{1}\{x > z\},$$

where $a \geq 0$ is a baseline probability of private-school attendance, $z \in (0, 1)$ is an ability threshold, and $\beta > 0$ captures cream skimming, with $a + \beta \leq 1$ so that attendance probabilities are well defined and the implied public-school density below remains nonnegative. Students above the threshold are therefore more likely to attend private school.

The aggregate private-school share is

$$s = \Pr(\text{private}) = \int_0^1 \Pr(\text{private} \mid X = x)\, dx = a + \beta(1 - z).$$

Thus a is the baseline attendance probability, while s is the actual private share once cream skimming is taken into account.

Conditional on attending private school, the density of X is

$$f_{X|\text{priv}}(x) = \begin{cases} \dfrac{a}{s}, & x \leq z, \\[2ex] \dfrac{a + \beta}{s}, & x > z. \end{cases}$$

Similarly, conditional on attending public school,

$$f_{X|\text{pub}}(x) = \begin{cases} \dfrac{1-a}{1-s}, & x \le z, \\ \dfrac{1-a-\beta}{1-s}, & x > z. \end{cases}$$

These expressions imply the sector means

$$\mu_{\text{priv}}(s) = E[X \mid \text{private}] = \frac{1}{2} + \frac{\beta z(1-z)}{2s},$$

and

$$\mu_{\text{pub}}(s) = E[X \mid \text{public}] = \frac{1}{2} - \frac{\beta z(1-z)}{2(1-s)}.$$

Three implications follow immediately.

First, the private-school mean exceeds the public-school mean whenever $\beta > 0$. This gap arises purely from selection, not from any productivity effect of private schools.

Second, the population mean remains constant:

$$s\,\mu_{\text{priv}}(s) + (1-s)\,\mu_{\text{pub}}(s) = \frac{1}{2}.$$

As students are reallocated across sectors, average performance in the population does not change.

Third, both sector means decline as private share rises. The private mean falls because, as private enrollment expands, there is less cream left to skim. The public mean also falls because a larger share of high-ability students has already sorted into the private sector.

Figure 1 displays this pattern with one modification made for visual display. With the over-draw factor β held fixed, even a vanishingly small private sector skims a finite amount of cream, so the public mean sits strictly below the population mean as $s \to 0$ and the private mean strictly above it as $s \to 1$. The figure instead draws a mean-preserving variant in which the skimming intensity fades as either sector absorbs the whole population—private premium proportional to $\sqrt{(1-s)/s}$ and public deficit proportional to $\sqrt{s/(1-s)}$—so that the public mean meets the population mean at $s = 0$ and the private mean meets it at $s = 1$. The population mean is exactly constant under either version, and the qualitative signatures are identical.

Commune-Level Market Dynamics in the Mature Period

Appendix Tables B1 and B2 summarize the descriptive evidence on commune-level reallocation in the post-2005 period using the balanced 4B sample, which makes 2005 and 2024 directly comparable. Even though changes between successive observed SIMCE waves are usually small, cumulative reallocation over 2005–2024 is not: the median commune's private share rises by 6.8 percentage points, the interquartile range runs from 0 to 16.9 points, and the full range spans roughly -30 to $+54$ points. Of the 333 communes observed at both endpoints, 224 see a higher private share in 2024, 41 see a lower share, and 68 show no net change. In the same balanced sample, the enrollment-weighted national private share rises from 45.9 percent in 2005 to 58.8 percent in 2024.

The correlates suggest that demographic contraction is part of the backdrop but not the whole story. Private-share growth is only weakly related to total enrollment change, but it is much more tightly linked to sector-specific adjustment, especially public-sector decline. This pattern is consistent with a mature market in which shrinking cohorts are absorbed unevenly across sectors rather than proportionally.

Sector and Market Definition Robustness

Appendix Table C1 re-estimates the baseline panel and cohort value-added specifications under alternative treatments of the fully private (PP) sector: including PP schools in scores and defining the private share as (PS+PP)/total with no commune exclusions, and keeping the PS-only definitions while retaining the high-

Table B1—Commune-Level Market Dynamics in the Balanced 4B Sample

	Mean	P25	P50	P75	Min	Max	N
Change in private share between observed SIMCE waves (pp)	0.56	-0.68	0.00	2.01	-22.67	39.68	5338
Change in private share, 2005–2024 (pp)	9.04	0.00	6.77	16.85	-29.95	54.43	333
Total enrollment change, 2005–2024 (%)	-6.03	-23.85	-10.81	1.90	-77.78	169.12	333
Public enrollment change, 2005–2024 (%)	-19.04	-37.50	-24.75	-8.51	-78.60	170.11	333
Private enrollment change, 2005–2024 (%)	36.66	-8.70	13.49	52.17	-100.00	825.00	249

Source: Author's calculations from the balanced 4B commune panel. The first row reports changes between successive observed SIMCE waves; the remaining rows compare commune endpoints in 2005 and 2024. The private-enrollment row excludes communes with zero private enrollment in 2005.

Table B2—Correlates of Commune-Level Private-Share Change

Variable	Correlation with Δ private share	N
Baseline private share (2005)	0.074	333
Baseline SES mean (2005)	0.174	333
Baseline low-SES share (2005)	-0.140	333
Total enrollment change, 2005–2024 (%)	0.096	333
Private enrollment change, 2005–2024 (%)	0.505	249
Public enrollment change, 2005–2024 (%)	-0.465	333

Source: Author's calculations from the balanced 4B commune panel. Correlations are unweighted across communes with both endpoints observed. Positive values indicate that communes with larger values experienced larger increases in private share over 2005–2024.

PP communes. The baseline exclusions are conservative—including the PP sector raises both estimates (panel: 12.2 vs. 10.5; cohort VA: 17.8 vs. 14.1).

Table C1—Robustness: Sector Definition and Sample Selection

	Panel FE			Cohort VA		
	(1) Baseline	(2) PS+PP	(3) No excl.	(4) Baseline	(5) PS+PP	(6) No excl.
Private share	10.52***	12.16***	10.43***	14.05***	17.78***	13.16***
	(2.90)	(2.84)	(2.82)	(2.47)	(2.38)	(2.41)
Private universe	PS only	PS+PP	PS only	PS only	PS+PP	PS only
High-PP excluded	Yes	No	No	Yes	No	No
Observations	19,221	19,821	19,815	6,142	6,342	6,342
Communes	336	346	346	315	325	325

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table C2 varies the market definition: aggregating Greater Santiago (the Provincia de Santiago plus Puente Alto and San Bernardo) into a single market, as Hsieh and Urquiola (2006) did, and excluding the Metropolitan Region entirely. The baseline panel estimate is stable (8.4 with Greater Santiago aggregated and 10.3 with the Metropolitan Region excluded, against 10.5), and the cohort value-added estimate rises under both variations (16.6 and 17.7, against 14.1), consistent with the cohort-fidelity evidence that commune-level markets are cleanest outside the capital. The commune × grade panel estimate, in contrast, loses size and significance under the same variations (2.3 and 1.8, SEs of roughly 2.4–2.7, not tabulated): with commune × grade fixed effects, the short-run temporal association is concentrated in the Greater Santiago market. The 10th-grade designs that carry the paper's main conclusions do not share this sensitivity.

Additional Cohort Value-Added Results

Appendix Table D1 reports the cohort-specific value-added estimates plotted in Figure 8. Appendix Figure D1 shows the corresponding subject-specific estimates.

Table C2—Robustness: Market Definition

	Panel FE			Cohort VA	
	(1)	(2)	(3)	(4)	(5)
	Communes	GS one mkt.	Excl. RM	GS one mkt.	Excl. RM
Private share	10.52***	8.41***	10.32***	16.62***	17.72***
	(2.90)	(2.81)	(3.14)	(2.42)	(2.79)
Observations	19,221	17,601	16,523	5,602	5,242
Markets	336	309	291	288	270

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table D2 reports the cohort-fidelity robustness discussed in Section VI: the baseline value-added estimate re-estimated on subsamples of commune-cohort cells whose 2M-to-4B tested-count ratios indicate high student overlap.

Table D1—Cohort-Specific Value-Added Estimates

4B year	2M year	PS-share coef.	SE	95% CI	Communes
2006	2012	12.92	(3.15)	[6.73, 19.12]	304
2007	2013	12.33	(3.71)	[5.03, 19.62]	304
2008	2014	10.76	(3.52)	[3.84, 17.68]	298
2009	2015	17.80	(3.05)	[11.81, 23.79]	305
2010	2016	15.42	(3.02)	[9.48, 21.37]	307
2011	2017	17.99	(3.15)	[11.80, 24.18]	309
2012	2018	14.22	(2.86)	[8.59, 19.85]	309
2016	2022	11.24	(1.89)	[7.53, 14.95]	312
2017	2023	12.32	(2.34)	[7.72, 16.92]	311
2018	2024	15.28	(2.66)	[10.05, 20.51]	314

Each row is a separate regression of the commune 2M score on the 2M PS share and the same cohort's 4B score, pooling subjects with subject fixed effects, weighted by 2M tested students, with standard errors clustered by commune. A joint test that the PS-share slope is equal across cohorts, from a fully interacted specification, gives $p = 0.002$.

Appendix Table D3 reports the flexible-baseline robustness discussed in Section VI. Each column re-estimates the value-added specification with the 4B control entering as a progressively more flexible function, specific to each subject × cohort cell; the baseline score is standardized within cell so that the nonlinear terms can-

Table D2—Cohort Value-Added: Robustness to Cohort Fidelity

	(1) All cells	(2) Drop heavy in/out	(3) Strict band
Private Share (2M)	14.05***	15.47***	16.79***
	(2.47)	(2.13)	(2.59)
Baseline Score (4B)	0.73***	0.73***	0.70***
	(0.07)	(0.05)	(0.06)
Sample	All cells	Ratio 0.5-1.5	Ratio 0.75-1.25
Observations	6,142	4,927	2,993
Communes	315	299	246
Adj. R-sq	0.511	0.548	0.591

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

not pick up cross-cohort scale differences otherwise absorbed by the fixed effects. The private-share coefficient ranges from 11.9 to 14.0 and remains significant at the 1 percent level in every specification.

For the *linear* adjustment, the insensitivity can be stated exactly. Conditional on the sample, weights, and fixed effects, imposing a persistence coefficient λ on the 4B score amounts to regressing $Y^{2M} - \lambda Y^{4B}$ on the private share, so the private-share estimate is an affine function of the imposed coefficient: $\hat{\beta}(\lambda) = 15.93 - 2.56\,\lambda$. The slope is small because, conditional on the subject $\times$ cohort fixed effects, the private share is only weakly correlated with the baseline (0.26 points per 10 percentage points of share). The freely estimated $\lambda = 0.73$ gives the baseline estimate of 14.05; unit persistence—a gain-score regression—gives 13.4; and even the wide range $\lambda \in [0.5, 1.2]$ implies estimates between 12.9 and 14.7. What the affine argument leaves open is nonlinearity in the achievement-persistence relationship, which is exactly what the flexible specifications in the table address.

Additional SES Evidence

Appendix Table E1 reports the gap-on-gap regressions discussed in the main text. It formalizes the relationship in Figure 7 by regressing the private-public score gap

Table D3—Cohort Value-Added: Flexible Baseline Adjustment

	(1) Linear	(2) Cell slopes	(3) Quadratic	(4) Spline	(5) Quintiles
PS share, grade 10	14.05***	13.77***	13.13***	13.03***	11.85***
	(2.47)	(2.48)	(2.46)	(2.44)	(2.61)
Observations	6,142	6,142	6,142	6,142	6,142
Communes	315	315	315	315	315

Standard errors in parentheses
Each column re-estimates the baseline value-added specification with the 4B control entering as the indicated function of the baseline score, standardized within each subject × cohort cell: pooled linear slope (column 1, the paper baseline), cell-specific linear slopes, quadratics, restricted cubic splines (4 knots), and quintile indicators. Weighted by 2M tested students. Standard errors clustered by commune. Common sample.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

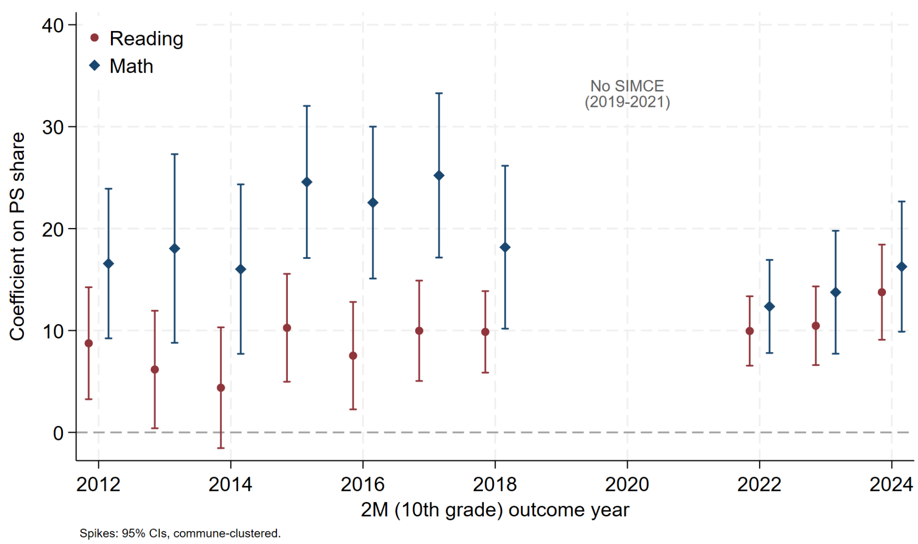

Figure D1. Cohort-specific value-added estimates by subject. Each point is a separate regression of commune 2M scores on the 2M PS share and the same cohort's 4B scores for a single subject, weighted by 2M tested students. Spikes are 95% confidence intervals based on commune-clustered standard errors.

on the SES gap, with and without conditioning on private share.

Table E1—Private-Public Score Gap on SES Gap

	(1) ΔY on ΔS	(2) + Priv Share	(3) No Commune FE
delta_ses	18.182***	17.734***	22.704***
	(1.839)	(1.569)	(1.587)
share_priv		7.809	9.341***
		(6.888)	(3.517)
Commune FE	Yes	Yes	No
GxSxT FE	Yes	Yes	Yes
Weight	Harmonic	Harmonic	Harmonic
N	13760	13760	13760
R2	0.637	0.638	0.466

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table E2 reports the sector-specific regressions discussed in Section V: within commune, sector mean scores decline with private share, and decline more steeply where sector SES separation is large, while the population slope remains positive and flat in ΔS.

Appendix Table E3 reports the robustness check that replaces the scalar SES-gap control with categorical SES-composition gaps and their interactions with private share.

Appendix Table E4 extends the SES-gap interaction test of equation (4) to the cohort value-added design, which conditions on lagged 4th-grade scores for the same cohort. On the common-support sample the baseline value-added estimate is $\hat{\beta} = 13.08$ ($p < 0.001$); adding the contemporaneous SES-gap interaction leaves the slope at the mean gap essentially unchanged ($\hat{\beta} = 13.04$), with a positive but imprecisely estimated interaction ($\hat{\theta} = 6.19$). The cohort design thus matches the panel result: the scope for socioeconomic sorting does not explain the private-share association.

Table E2—Sector-Specific Outcomes with SES Gap Interaction

	(1) Pub: Base	(2) Pub: Intxn	(3) Priv: Base	(4) Priv: Intxn
share_priv	-12.319**	-13.010**	-3.902	-4.893
	(5.194)	(5.042)	(3.520)	(3.483)
delta_ses_c	-7.432***	0.332	7.166***	13.317***
	(1.334)	(2.514)	(0.963)	(1.801)
p_x_dses		-19.011***		-10.346***
		(4.501)		(3.655)
Commune FE	Yes	Yes	Yes	Yes
GxSxT FE	Yes	Yes	Yes	Yes
DV Sector	Public	Public	Private	Private
N	13760	13760	13760	13760
R2	0.697	0.705	0.730	0.733

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure E1 shows that the SES gap between sectors has been remarkably stable over time, with public schools averaging around 2.0 on SIMCE's 5-point SES scale, PS schools around 3.0, and PP schools around 4.5.

Figures E2 and E3 show the distribution of school-level scores by SES group and sector. Within each SES group, private schools score higher than public schools of the same SES category—descriptive support for a productivity advantage, though the coarse categories leave room for sorting within them.

Table E3—Robustness: Categorical SES Composition Gaps

	(1) Cat Controls	(2) Cat Interactions
share_priv	8.738***	8.656*
	(2.997)	(4.479)
delta_sh_ses1	-0.189	17.132
	(8.394)	(18.350)
delta_sh_ses2	-1.580	12.777
	(8.295)	(18.233)
delta_sh_ses3	0.022	14.843
	(7.900)	(17.753)
delta_sh_ses4	0.600	18.932
	(8.483)	(17.850)
p_x_dsh1		-32.945
		(44.315)
p_x_dsh2		-27.570
		(44.115)
p_x_dsh3		-27.983
		(42.985)
p_x_dsh4		-35.851
		(42.774)
Commune FE	Yes	Yes
GxSxT FE	Yes	Yes
N	13760	13760
R2	0.756	0.757

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E4—Cohort Value-Added with SES Gap Interaction

	(1) Baseline VA	(2) SES Interaction
Private Share (10th Grade)	13.079***	13.043***
	(2.916)	(2.855)
Baseline Score (4th Grade)	0.726***	0.731***
	(0.076)	(0.071)
SES Gap (Centered)		-2.294
		(2.122)
Private Share x SES Gap		6.187
		(4.455)
Subject x Cohort FE	Yes	Yes
N	4272	4272
Clusters	222	222
R2	0.531	0.536

Sample: Common-support matched 4B-to-2M cohorts with observed sector-specific SES measures in 2M.
Unit of observation: commune × subject × cohort.
Weights: number of tested students in 2M.
Fixed effects and clusters are reported in the table. Standard errors are clustered by commune.

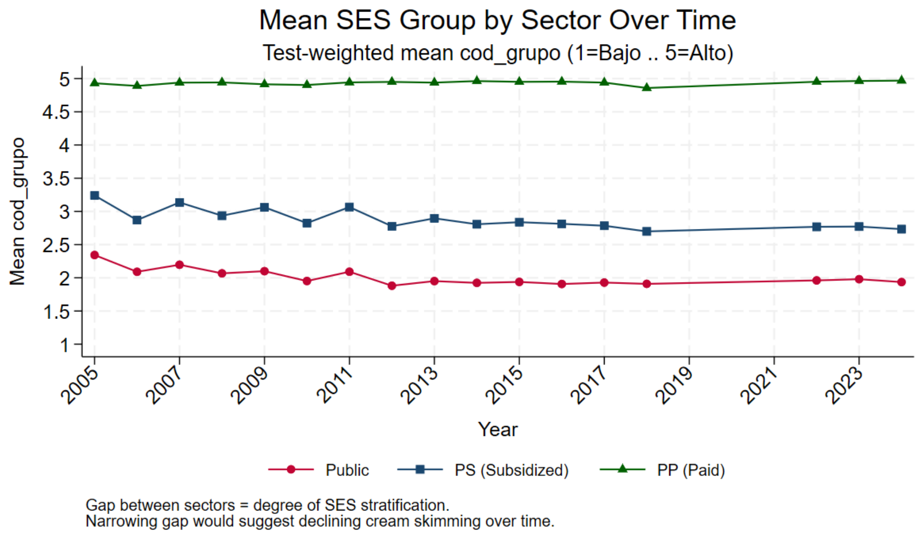

FIGURE E1. MEAN SCHOOL-LEVEL SES GROUP BY SECTOR OVER TIME. THE GAP BETWEEN SECTORS IS STABLE ACROSS THE SAMPLE PERIOD.

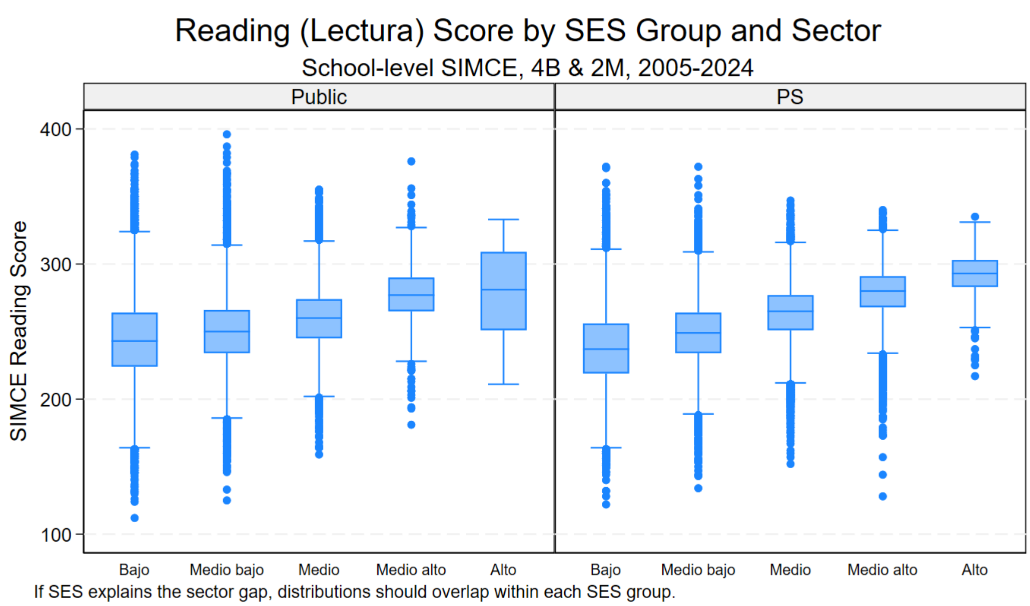

FIGURE E2. READING SCORE DISTRIBUTIONS BY SES GROUP AND SECTOR. WITHIN SES GROUPS, PRIVATE SCHOOLS TEND TO OUTPERFORM PUBLIC SCHOOLS.

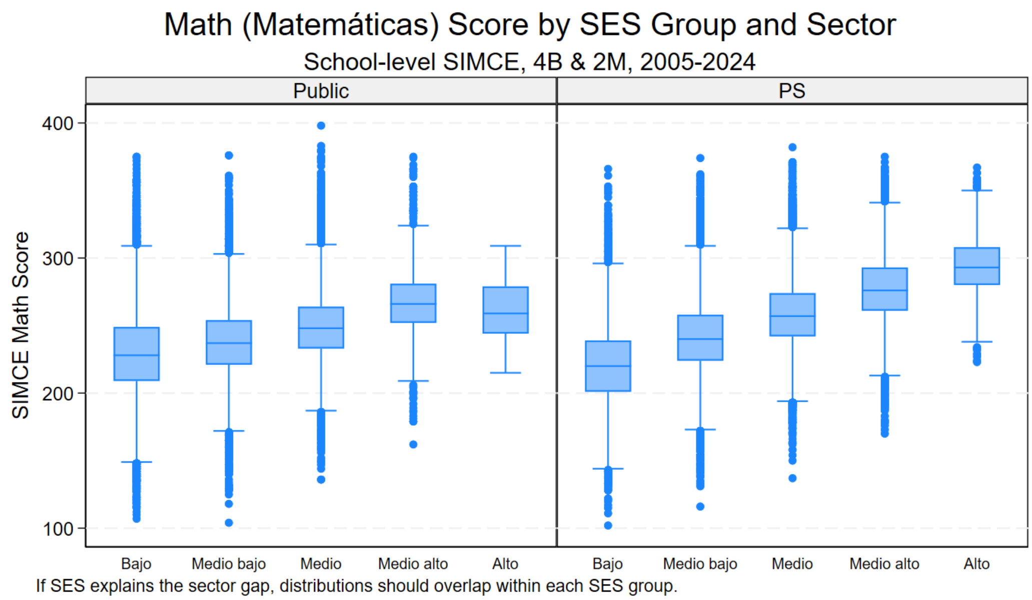

Figure E3. Math score distributions by SES group and sector.

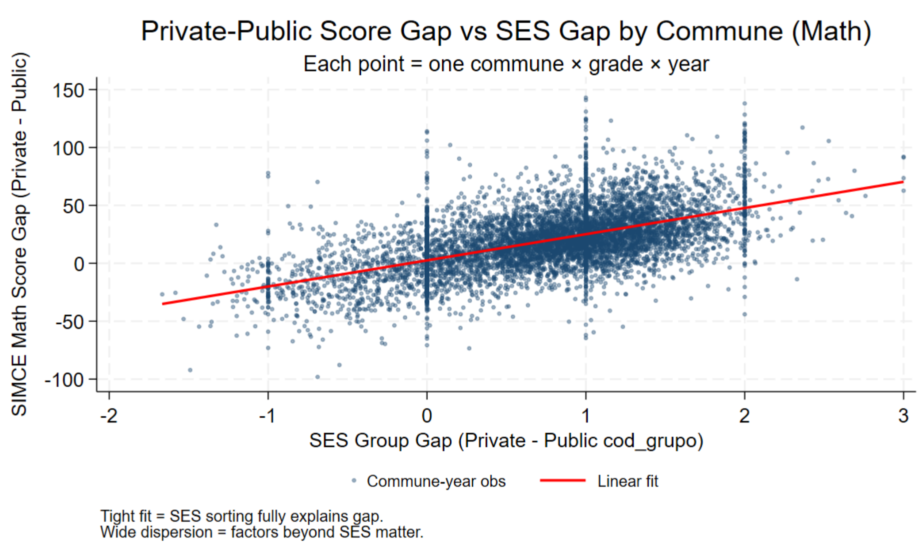

Figure E4. Private-public score gap vs. SES gap by commune-year, math. Construction is identical to Figure 7 using math scores and math tested-student weights.