Teacher Preparation, Pre-College Human Capital and Student Learning: Evidence from *Enseña Chile*

Soledad De Gregorio[a], Christopher A. Neilson[b,*], Sebastián Gallegos[c]

[a] *University of Southern California, Los Angeles, USA*
[b] *Yale University, New Haven, USA*
[c] *Universidad Adolfo Ibáñez, Santiago, Chile*

Abstract

Teacher value-added is rarely estimated in middle-income countries, which lack the repeated classroom-level testing it requires. We administered baseline and end-of-year mathematics tests in Chilean classrooms and linked the resulting estimates to teachers' pre-college admission scores, experience, and preparation route. We compare *Enseña Chile*, the Chilean Teach For All affiliate, with traditional teacher-college preparation. A one-SD (σ) higher admission score predicts 0.41σ higher value-added. *Enseña Chile* teachers gain $+0.25\sigma$ from ability but trail traditional teachers by 0.16σ after reweighting, driven by inexperience. Training routes are statistically indistinguishable. High-ability recruits help only if they stay long enough to gain experience.

Keywords: teacher value-added, alternative certification, teacher selection, Chile, Teach For All
JEL: I20, I21, I28, J45

---

*Corresponding author.

Email addresses: soleca@gmail.com (Soledad De Gregorio), christopher.neilson@yale.edu (Christopher A. Neilson), sebastian.gallegos@uai.cl (Sebastián Gallegos)

1. Introduction

High-quality teachers are among the most important school-based inputs for student achievement and long-run outcomes (Hanushek and Rivkin, 2010; Jackson et al., 2014). Yet beyond well-established returns to experience, observable teacher traits correlate weakly and inconsistently with effectiveness, fueling persistent debate over how to prepare and select teachers. In parallel, many school systems face acute staffing shortages in specific subjects and geographies. Alternative entry pathways, most prominently Teach For America (TFA) and the Teach For All network of country affiliates, aim to attract high-ability graduates to teach in hard-to-staff schools through short, intensive pre-service training followed by structured in-service coaching.

A central question in this literature is *why* TFA-style programs achieve roughly comparable test-score results to traditional preparation, despite radically shorter coursework. Two mechanisms are typically invoked: *selection* on observable cognitive ability and motivation, and *intensive coaching* during the two-year commitment. Distinguishing the two has been difficult because almost no study has both individual-level pre-college ability scores for the relevant teachers *and* student-level panel data sufficient to estimate value-added (VA). Consequently, the literature documents the average result but leaves its composition implicit.

This paper provides such a decomposition. We study *Enseña Chile* (eCh), the Chilean Teach For All affiliate. We link administrative records on teachers' college entrance exam scores (PSU/PAA) and years of experience to research-administered beginning- and end-of-year SEPA[1] mathematics assessments for grades 9–11 across *Enseña Chile* partner schools (those that host eCh teachers) and non-partner comparison schools (which do not). The resulting dataset—114 mathematics teachers (37 eCh, 77 traditional), 169 classes, 3,756 students, 59 schools—lets us measure each teacher's VA and decompose the *Enseña Chile*–traditional gap into pre-college ability, experience, and a route residual that absorbs training, motivation, unobserved selection, and any ability-by-experience interaction.

We report three findings. First, raw VA for *Enseña Chile* teachers is 0.32σ below that of traditional teachers (-0.19 vs. $+0.13$); among the 104 teachers with observed entrance-exam scores, reweighting the comparison group to the national common-support PSU/PAA distribution narrows the gap from -0.26σ to -0.16σ. The gap largely reflects inexperience: all 37 *Enseña Chile* teachers are in their first year (20) or second (17). Second, *Enseña Chile* teachers are positively selected on ability: a one-SD increase in PSU/PAA is associated with a 0.41σ increase in VA, and *Enseña Chile* teachers outscore traditional teachers in partner schools by 64 points in math and 114 in language (close to one SD of the national teacher distribution on the combined measure). Yet conditioning on this ability advantage alone *widens* the raw gap rather than closing it.

To separate the remainder into experience and route of preparation, we draw on administrative records showing that about half the comparison teachers are themselves early-career, which gives genuine overlap with the

[1] *Sistema de Evaluación de Progreso del Aprendizaje*; details in Section 2.

Enseña Chile cohort. The route residual is statistically indistinguishable from zero, though imprecise: second-year *Enseña Chile* teachers reach parity with comparably early-career traditional peers. Third, within-*Enseña Chile* learning is rapid: the second-year cohort displays VA 0.60σ above the first-year cohort, consistent with steep early-career returns and the program's structured coaching model.

We explicitly distinguish between two comparisons. The *staffing* comparison (without ability and experience controls) is the one school leaders face when filling hard-to-staff vacancies; under it, *Enseña Chile* teachers reach rough parity with the traditional teachers already in their schools. The *conditional* comparison (adding PSU/PAA and experience) is the regression-estimated object and isolates the route-of-preparation residual, which is statistically indistinguishable from zero, though imprecise. Any staffing recommendation is an interpretation rather than an estimate, since we never observe the marginal teacher a vacancy would otherwise draw; we return to this distinction in Section 4. Subject to that caveat, *Enseña Chile*'s strong positive selection makes it a plausible lever for raising instructional quality where able candidates are scarce.

This paper makes three contributions. The first establishes the evidence base: a purpose-designed linked dataset built to identify the PSU/PAA–VA gradient. The remaining two use that evidence to speak to teacher selection and preparation.

First, we build a linked dataset for a setting where teacher VA measures are scarce. Estimates remain rare in middle-income countries because the longitudinal, classroom-level testing they require is seldom available. We administered baseline and end-of-year SEPA mathematics assessments across 169 grade 9–11 classrooms and linked the resulting VA estimates to administrative PSU/PAA entrance-exam scores and the Ministry of Education's teacher experience panel. A representative traditional comparison would conflate route with ability: *Enseña Chile* recruits are drawn from selective universities and concentrate at the top of the PSU/PAA distribution. We instead stratified non-partner schools by their teachers' PSU/PAA decile, inviting at least one school per decile and trading representativeness for the within-comparison ability overlap needed to estimate the PSU/PAA–VA gradient. For Latin America, the leading individual-teacher-effects evidence comes from randomized classroom assignment in Ecuador (Araujo et al., 2016); our data extend that evidence base to Chile with a different design.

Second, we estimate how teachers' pre-college academic ability predicts VA. To our knowledge, this is the first individual-teacher-level VA estimate using a country's selective-university entrance exam, taken before teacher preparation, as the focal cognitive measure. Prior work instead uses post-training subject knowledge (Hill et al., 2005; Metzler and Woessmann, 2012; Bietenbeck et al., 2018), licensure tests (Clotfelter et al., 2007), pre-hire screening batteries (Rockoff et al., 2011; Jacob et al., 2018), or country-level PIAAC scores (Hanushek et al., 2019). The closest pre-college precedent is Boyd et al. (2008), where SAT scores and college selectivity enter as controls among many rather than as the focal measure. Our companion study links pre-college achievement to broader Chilean teacher outcomes (Neilson et al., 2023); here we isolate the VA margin and the alternative-pathway comparison.

Third, we decompose selection and experience for a Teach For All affiliate at the individual-teacher level. Prior TFA evidence, mostly from the United States, asks whether alternative-pathway teachers are more or less effective than traditional teachers and generally finds parity or modest math gains (Glazerman et al., 2006; Xu et al., 2011; Chiang et al., 2017; Clark et al., 2017; Clark and Isenberg, 2020; Backes and Hansen, 2024), a pattern a recent meta-analysis of TFA impacts confirms (Citkowicz et al., 2024). Evidence from England and Peru points in a similar direction for Teach For All affiliates outside the United States (Allen and Allnutt, 2017; Lavado and Guzmán, 2020). What remains unclear is whether these average effects reflect who enters the pathway, what they learn on the job, or the route itself. By linking PSU/PAA scores, experience, and SEPA-based VA, we can separate these channels directly.

The remainder of the paper proceeds as follows. Section 2 describes the data, sample, VA construction, and decomposition approach. Section 3 presents the main results. Section 4 concludes by discussing interpretation, policy counterfactuals, and limitations.

2. Data and Methods

To our knowledge, no existing Chilean or Latin American source links research-administered achievement gains, student perception surveys, and administrative ability and experience records for the same teachers. The analytic dataset combines a student testing campaign with two administrative sources. We administered the SEPA mathematics assessment at the beginning (May) and end (November) of the 2016 school year in grades 9–11 through external proctors.[2] The baseline covers prior-grade content and the follow-up current-grade content. Alongside the November tests, we administered a Teach For All student perception survey aligned with the Measures of Effective Teaching (MET) study domains (rigorous expectations, engagement, supportive relationships). Those sources supply PSU/PAA math and language scores and, separately, years of experience from the MINEDUC teacher administrative panel.[3] Our ability measure is the simple average of a teacher's mathematics and language scores. Teachers who sat the pre-2004 *Prueba de Aptitud Académica* and those who sat its successor, the *Prueba de Selección Universitaria*, are pooled; both are normalized to the same mean and standard deviation, so scores from the two regimes are on a common scale. Among all test-takers, the entrance exam has mean 500 and SD 110; teachers occupy a narrower range, where the combined math-language SD is 80 points. Every ability-in-SD statement in the paper uses this teacher-population SD.

The working sample comprises 3,756 students in 169 classes taught by 114 mathematics teachers across 59 schools. It includes 37 *Enseña Chile* teachers in eCh partner schools, 31 traditional teachers in those same partner schools, and 46 traditional teachers in 24 non-partner schools from the Metropolitan Region. Of the 53 eligible eCh partner schools, 38 (72%) participated; 35 contribute teachers to the analytic VA sample. Table 1 reports descriptive statistics by teacher group, including teacher characteristics, school covariates, SEPA scores, and observation counts.

[2] SEPA was developed by MIDE-UC, the measurement center at the *Pontificia Universidad Católica de Chile*.

[3] Formally the *Docentes* (*Idoneidad Docente*) panel.

The comparison sample is purposive. Panel A of Table 1 shows that *Enseña Chile* teachers have substantially higher PSU/PAA scores than traditional teachers in partner schools, while the traditional comparison teachers provide overlap across the lower and upper parts of the PSU/PAA distribution; Panel C reports baseline and follow-up SEPA scores by group. We therefore stratified non-partner schools by their teachers' PSU/PAA decile and invited at least one school per decile. Participating non-partner schools cover deciles 1–10, placing high- and low-PSU/PAA traditional teachers alongside the *Enseña Chile* group (see Online Appendix A for details). The design therefore trades representativeness for the within-comparison ability overlap needed to estimate the gradient. Accordingly, the unweighted group means and gaps are analytic-sample comparisons rather than population mean VA estimates for Chilean teachers.

The empirical strategy has three steps. First, following Chetty et al. (2014), we construct teacher VA from the 2016 SEPA gain design, using those authors' `vam` implementation. Scores are residualized on school and class covariates—region, municipal dependency, school SES tier, and class size—with teacher effects estimated from within-teacher variation across the baseline and follow-up waves; class-level follow-up residuals are then projected on baseline class residuals, and the predicted residual learning is standardized across teachers in the analytic sample. Teacher effects are reported unshrunk, so one σ of VA is one standard deviation of the *estimated* teacher-effect distribution in this sample. With a single year of data those estimates carry classroom-level noise, which inflates that standard deviation and so attenuates the standardized gradients reported below.

Table 1: Summary statistics by teacher group.

	***Enseña Chile* Teachers**	**Traditional (eCh schools)**	**Traditional (non-eCh)**	**Total**
Panel A: Teacher Characteristics				
Experience (years)	0.5	8.0	14.0	8.6
	(0.5)	(9.6)	(12.1)	(11.2)
Language PSU/PAA	640	526	565	584
	(58)	(73)	(74)	(81)
Math PSU/PAA	680	616	629	644
	(84)	(79)	(75)	(83)
Panel B: School Covariates				
Metropolitan Region	0.31	0.41	1.00	0.67
Public School	0.28	0.29	0.12	0.20
Low SES School	0.49	0.32	0.18	0.30
Mid-Low SES School	0.46	0.57	0.32	0.41
Mid SES School	0.05	0.05	0.19	0.12
Mid-High SES School	0.00	0.05	0.31	0.16
High SES School	0.00	0.00	0.00	0.00
Class Size	21.2	23.3	29.0	25.5
	(7.2)	(6.5)	(7.1)	(7.9)
Panel C: SEPA Scores				
Baseline	627	639	650	641
	(35)	(37)	(41)	(40)
Follow-up	634	643	655	647
	(31)	(35)	(40)	(37)
Gain	7.2	4.6	5.6	5.9
	(28.4)	(29.7)	(27.4)	(28.2)
Panel D: Observations				
N Students	1,155	754	1,847	3,756
N Classes	64	36	69	169
N Teachers	37	31	46	114
N Schools	30	21	24	59

Notes: Entries are student-level means, with standard deviations in parentheses where applicable. SES categories follow the SIMCE socioeconomic classification. Teacher experience is measured as years of service in the MINEDUC *Docentes* panel. About half of the comparison teachers have five or fewer years of experience.

Second, we assign each student the VA estimate of their mathematics teacher and estimate pooled regressions of the form

$$\widehat{\text{VA}}_{ij} = \alpha + R_j'\theta + \beta \text{PSU}_j + \phi_1 \text{Exp}_j + \phi_2 \text{Exp}_j^2 + X_{ij}'\gamma + \varepsilon_{ij},$$

where i indexes students, j indexes teachers, R_j indicates route/school group, and X_{ij} includes SES tier, grade fixed effects, and class size. Baseline SEPA enters through the constructed VA measure rather than as a separate second-stage control. Model 1 omits PSU/PAA and experience, corresponding to the staffing comparison; Model 2 adds them, corresponding to the conditional comparison; column 3 adds the leave-out school faculty-composition control. We do not estimate school fixed effects because the within-school cells are too thin, with about 1.9 sampled teachers per school (114 teachers across 59 schools). Standard errors are clustered at the classroom level. Models 1 and 2 are estimated on the 104 teachers (3,454 students) with observed PSU/PAA scores, so the staffing and conditional comparisons use a constant sample; column 3 drops a further 56 students (see the Table 2 note).

Third, we decompose the reweighted *Enseña Chile*–traditional gap into selection on PSU/PAA, early-career experience, and a route residual. This step uses the corrected experience records together with a separate teacher-level weighted regression of VA on route, PSU/PAA (per 100 points), and a first-year indicator, a specification distinct from the student-level Model 2 of Table 2. Experience identifies the early-career overlap that underlies the experience and route components, but an earlier administrative file mismeasured it for many comparison teachers; we detail the correction below. For this exercise, the *Enseña Chile* cohort mean remains unweighted and the sampled comparison teachers are reweighted to the relevant national teacher PSU/PAA distribution over common-support deciles (Online Appendix Section B.8).[4]

Given the purposive design, this empirical exercise is a within-sample VA comparison rather than a population estimate. We estimate VA as cross-

[4]Online Appendix B reports robustness checks on each of these choices.

sectional teacher fixed effects from a single school year, so we do not exploit within-teacher variation across cohorts and cannot rule out classroom-level unobservables. Three features mitigate this concern. First, the VA construction conditions on baseline achievement, absorbing prior achievement differences within teacher. Second, although the design is too thin in within-school cells to support school fixed effects, we control directly for the academic composition of each school's teaching staff, holding constant the dimension of school-level sorting most likely to be correlated with teacher ability. Third, because the sample was built to span the teacher ability distribution, interpretation turns on within-comparison validity, through the baseline-score and faculty-composition controls just described, rather than on external representativeness.

Among eligible *Enseña Chile* partner schools, the reason recorded for non-participation is administrative wherever a reason is recorded at all: several already administered SEPA annually and could not re-test. For the remainder the project files record only that the school declined or did not respond, which is an outcome rather than a reason and so does not speak to quality-related selection in either direction (see Online Appendix A). We therefore cannot rule out selection on unobservables within the comparison. The direction of any resulting bias is ambiguous: improvement-oriented partner schools could pull *Enseña Chile* estimates upward, while less-needy schools that select out could pull them downward.

We measure experience from the MINEDUC teacher administrative panel rather than the earlier 2011 service file, which had mean-imputed about half the comparison teachers to mid-career. The corrected data show that 38 of 77 traditional teachers are themselves early-career (≤ 5 years of service; median three years in *Enseña Chile* partner schools), creating genuine overlap with *Enseña Chile* teachers in the experience distribution.[5] This overlap is what makes the experience and route components separable without the cohort-comparability assumption used in earlier drafts, since traditional teachers are now observed at the experience levels where *Enseña Chile* teachers sit. The decomposition itself conditions on a first-year indicator rather than matching on years of service, so its route residual compares *Enseña Chile* teachers with traditional teachers at all experience levels, holding PSU/PAA and first-year status fixed. Online Appendix B.1 reports the complementary check that drops first-year *Enseña Chile* teachers and compares second-year *Enseña Chile* teachers with traditional teachers directly, and the within-*Enseña Chile* first-to-second-year contrast provides a separate check on the first-year penalty.

The decomposition imposes an additively separable form: the modest teacher count and thin within-school cells do not identify an ability-by-experience interaction, which therefore loads onto the route residual. Two cautions also apply: no traditional teacher is observed in their literal first

[5] Of these 38, 31 have a year of service recovered directly from the *Docentes* panel (median 3, range 1–5). The remaining 7 are classified as early-career by their absence from the 2011 teacher census rather than by a measured year. That group is small but higher-performing (mean VA +0.74, versus −0.02 for the 31 panel-matched teachers), so the early-career traditional mean of +0.12 becomes −0.02 if they are dropped. None of the 7 has an entrance-exam score, so none of them enters the decomposition; the sensitivity affects the descriptive group means and Figure 2, not the components reported in Section 3.

year, so a generic first-year dip cannot be fully separated from an *Enseña Chile*-specific one; and the route residual itself is estimated from a small sample, so we report it with its confidence interval in Section 3.

3. Results

At the teacher level, *Enseña Chile* teachers' mean VA is -0.19σ versus $+0.13\sigma$ for traditional teachers, a gap of roughly 0.32σ (s.e. 0.16, $p = 0.04$). This raw gap is an analytic-sample contrast: the sampled traditional teachers were deliberately drawn to span the PSU/PAA distribution rather than to represent the national workforce. On the 104 teachers with observed entrance-exam scores it is -0.26σ, and reweighting the comparison teachers to the national common-support PSU/PAA distribution narrows it further to -0.16σ, with a 95% interval that includes zero ($[-0.36, +0.05]$; Online Appendix Table B.7). The reweighted gap is therefore not itself statistically distinguishable from zero. What the decomposition below identifies is its *composition*: as we show, the ability and inexperience components are individually well determined even though their net is imprecise, and the route residual is statistically indistinguishable from zero.

Table 2 reports the corresponding regression evidence. PSU/PAA strongly predicts VA: in the pooled Model 2, a 100-point higher entrance-exam score is associated with a 0.51σ increase in VA (classroom-clustered s.e. 0.12), significant at the 1% level and robust to teacher- and school-clustered standard errors. Scaled to the 80-point teacher-population SD, that is 0.41σ per SD of ability. The gradient is robust to the school faculty-composition control described below: it is essentially unchanged at 0.49σ per 100 points (classroom-clustered s.e. 0.12), or 0.39σ per SD, in column 3. Figure 1 visualizes this positive ability–VA relationship.[6]

The gradient is large, and the U.S. evidence it invites comparison with rests on different ability measures. That literature does find applicant cognitive and academic measures predictive of teacher performance: Rockoff et al. (2011) report moderately large associations between recruits' cognitive and non-cognitive factors and student achievement, and Jacob et al. (2018) find that academic background and screening scores strongly predict subsequent job performance. Two features of our measure plausibly bear on the magnitude. PSU/PAA is a high-stakes selective-university entrance exam taken before any teacher preparation, and it spans a wider ability range among Chilean teachers than the licensure tests and screening batteries used in U.S. studies, which are administered to an already-screened pool.

Consistent with this gradient, our companion study shows that pre-college academic achievement also predicts other long-run Chilean teacher outcomes: exit exams, in-classroom evaluations, wages, and employment (Neilson et al., 2023).

A remaining concern is that the teacher-level gradient may partly reflect where teachers work. To address this, we control for the academic composition of each school's teaching staff. For each sampled teacher, we compute the leave-

[6]Table 2 coefficients are student-weighted. The teacher-level fit behind the binned markers in Figure 1 instead weights teachers equally and partials out experience and route, yielding $+0.24\sigma$ per SD, a complementary summary. The analytic-sample and national-teacher PSU/PAA standard deviations are nearly identical (78 and 80 points), so the two figures differ by weighting and specification, not by scale.

Table 2: Regression models of teacher VA.

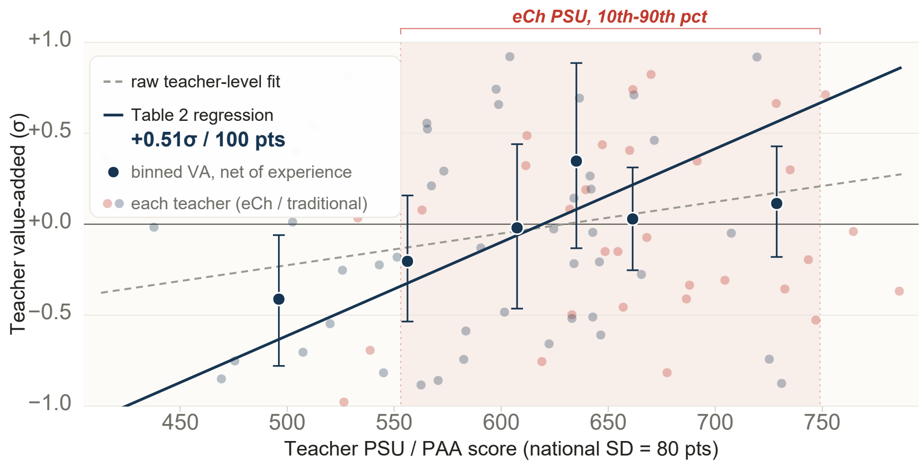
	(1) Model 1	(2) Model 2	(3) Model 2 + Faculty
Traditional teacher, partner school	0.275	0.457	0.422
	(0.234)	(0.241)	(0.234)
Enseña Chile teacher, partner school	0.160	0.116	-0.006
	(0.208)	(0.288)	(0.306)
PSU/PAA (per 100 pts)		0.514***	0.489***
		(0.123)	(0.119)
Experience (years)		0.039	0.019
		(0.033)	(0.037)
Experience^2		-0.0005	0.0000
		(0.001)	(0.001)
School faculty ability (std.)			0.207*
			(0.086)
Mid-Low SES	-0.343*	-0.404*	-0.342*
	(0.158)	(0.157)	(0.165)
Mid SES	-0.226	-0.350	-0.443*
	(0.229)	(0.235)	(0.218)
Mid-High SES	0.084	-0.248	-0.349
	(0.335)	(0.291)	(0.294)
Class Size	0.035**	0.042***	0.035**
	(0.013)	(0.010)	(0.012)
Grade	0.340**	0.332***	0.315***
	(0.109)	(0.096)	(0.094)
Constant	-4.254***	-7.642***	-7.038***
	(1.121)	(1.080)	(1.086)
R^2	0.121	0.283	0.312
Observations	3,454	3,454	3,398

Notes: Dependent variable is standardized teacher value-added assigned to students. The omitted route category is a traditional teacher in a non-partner school. The PAA/PSU coefficient is per 100 points of the entrance exam; one standard deviation of teacher PAA/PSU is about 80 points, so column 2 corresponds to 0.41σ per SD. SES tiers are the SIMCE groups, with Low omitted; no High-SES school appears in the sample. Model 2 adds PAA/PSU, experience, and experience squared; column 3 adds the leave-out school-faculty PAA/PSU percentile (the mean for the school's other teachers), which is undefined for teachers with no other PSU-linked teacher in their school; dropping those 56 students reduces the column 3 sample from 3,454 to 3,398, while columns 1 and 2 share the same 3,454-student sample. All models include SES tier, class size, and grade; classroom-clustered standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

out mean national entrance-exam (PSU/PAA) percentile of the school's *other* teachers, using the 2011 teacher census linked to DEMRE scores (available for 58 of our 59 schools). This continuous school-input measure is estimable across the sample, unlike school fixed effects, which the design cannot support.

Faculty composition is predictive in its own right: a one-SD increase is associated with $+0.21\sigma$ higher teacher VA (classroom-clustered s.e. 0.09;

Figure 1: Teacher-level value-added–PSU/PAA relationship.



Notes: Solid navy line: the student-level Model 2 PSU/PAA coefficient (Table 2). The $+0.51\sigma$ shown inside the figure is per 100 PSU/PAA points, matching the units of the horizontal axis; equivalently $+0.41\sigma$ per 80-point teacher-population SD. Dashed gray: unadjusted teacher-level fit. Shaded band: the 10th–90th percentile of the *Enseña Chile* PSU/PAA range. Faint dots: individual teachers; 25 of the 104 fall outside the plotted vertical range ($|\mathrm{VA}| > 1$) and are not drawn, though all 104 enter every fit and binned mean shown. Navy markers: PSU/PAA-binned VA, experience and route partialled out by Robinson-style partial-linear residualization; whiskers are teacher-bootstrap 95% intervals.

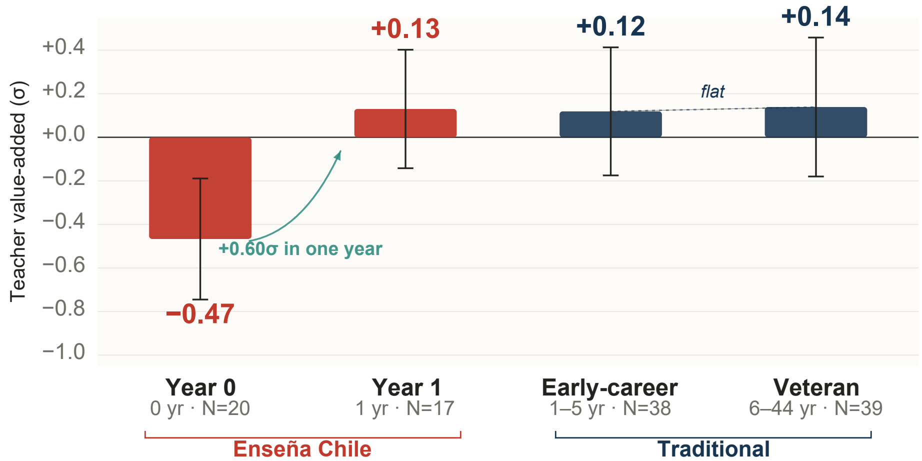
column 3 of Table 2), significant at the 5% level with classroom-clustered standard errors, though not at conventional levels under the more conservative teacher- and school-clustered alternatives ($p \approx 0.09$ and 0.12; Online Appendix Table B.6). Conditioning on it leaves the PSU/PAA gradient and the experience profile essentially unchanged: the gradient moves from 0.51σ to 0.49σ per 100 points, and the profile stays flat. The route terms are less stable. Measured against traditional teachers in non-partner schools, the *Enseña Chile* coefficient moves from +0.12 to −0.01, while its difference from the traditional-in-partner-school coefficient widens from −0.34 to −0.43. As discussed below, neither quantity separates route from school placement; the estimate that does, within schools, is +0.06 (s.e. 0.25; Online Appendix Section B.7). The ability finding therefore survives the placement control this design can support. That control is the limit of what the sample allows: at roughly two sampled teachers per school, individual ability and school-level sorting cannot be separated more finely, and the ability contribution is correspondingly an association within this sample rather than a causal return to ability. Online Appendix B.7 reports the school fixed-effect and between/within PSU/PAA decompositions that make this indeterminacy explicit.

Turning to experience, the pooled relationship is flat on average. The marginal-year coefficient is small (0.04, s.e. 0.03) and not statistically distinguishable from zero, consistent with Papay and Kraft (2015) and Ladd and Sorensen (2017) on diminishing returns past the early career.

The early-career pattern within *Enseña Chile* is much steeper. The within-route contrast captures the part of the experience profile most relevant for the program, since *Enseña Chile* teachers are concentrated in years 0–2. Second-year *Enseña Chile* teachers' VA averages $+0.13\sigma$, about 0.60σ above the first-year average of -0.47σ (Figure 2), a difference significant at the 1% level. This improvement is well above typical first-year returns documented in the U.S. literature. It should not be read as a pure return to experience, however, because it compares the 2015 and 2016 *Enseña Chile* cohorts (different teachers in a single testing wave) and therefore may also reflect cohort selection.

We next separate the selection and experience channels, using the experience-

Figure 2: Experience profile by route on the corrected experience data.



Notes: Mean VA by route and experience tier; whiskers are 95% intervals formed as the group mean ±1.96 standard errors. Experience corrected via the MINEDUC teacher administrative panel. The early-career traditional bar pools 31 teachers whose years are recovered from the panel with 7 bounded at ≤ 5 years by census non-appearance; see footnote 5. The *Enseña Chile* group means are reported in the body.

corrected data. Figure 3 decomposes the reweighted -0.16σ gap into three additive components: a $+0.25\sigma$ pre-college ability contribution from *Enseña Chile* teachers' PSU/PAA advantage, a -0.30σ first-year inexperience penalty, and a -0.10σ route-of-preparation residual (Online Appendix Section B.8).

The 37 *Enseña Chile* teachers are the cohort under evaluation rather than a draw from a wider pool, so the bootstrap resamples only the sampled traditional teachers and the component intervals are conditional on that cohort: $[+0.06, +0.42]$ for the ability contribution, which excludes zero, and $[-0.37, +0.17]$ for the route residual, which does not. This scheme yields no informative interval for the inexperience component: no sampled traditional teacher is in their first year, so its experience differential is the *Enseña Chile* cohort's own first-year share, held fixed by construction. An *Enseña Chile*-inclusive bootstrap that also resamples the cohort gives $[-0.55, -0.09]$, so the penalty remains clearly negative once cohort-sampling uncertainty is added, and leaves the other two conclusions unchanged (Online Appendix B.6).[7]

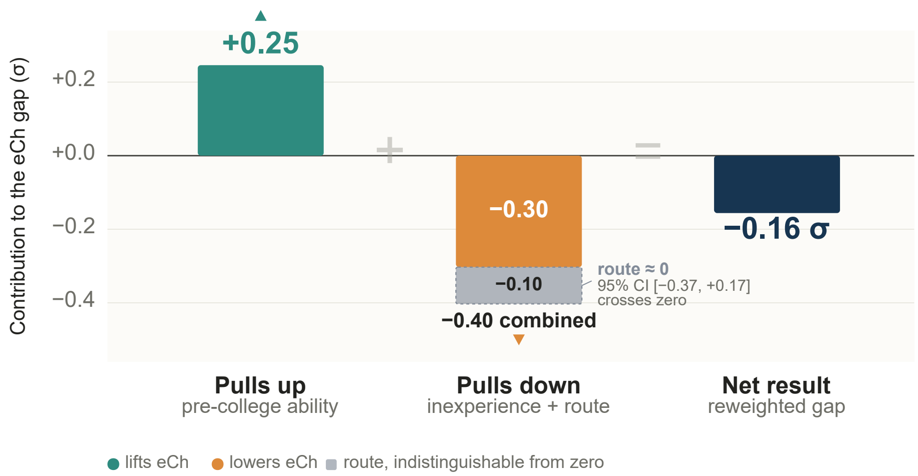
Consistent with a near-zero residual, second-year *Enseña Chile* teachers reach the traditional mean in the raw data ($+0.13\sigma$ for both groups), while the regression-adjusted *Enseña Chile* route coefficient in Table 2 is also near zero. The two route dummies are measured against traditional teachers in non-partner schools, so neither is by itself a route contrast. Their difference compares *Enseña Chile* teachers with traditional teachers in partner schools—-0.34 in Model 2, -0.12 without controls in Model 1—and mixes route with school placement. The estimate that separates the two is the within-school contrast, $+0.06$ (s.e. 0.25), reported in Online Appendix Section B.7. The residual is estimated from a modest number of teachers and remains imprecise: the data are consistent with pre-college ability plus rapid early-career learning fully accounting for the reweighted gap, but they cannot rule out a moderate route shortfall.[8]

Placing *Enseña Chile* teachers in the national distribution makes the selection-versus-experience tension concrete (Online Appendix Figure B.2). Using the fitted ability–experience model to predict VA for the PSU-era

[7] Online Appendix B.4 varies the assumed first-year penalty and traces the implied route residual.

[8] Conclusions survive dropping first-year *Enseña Chile* teachers (Model 1 *Enseña Chile* coefficient $+0.35$, s.e. 0.23; Online Appendix Table B.1), teacher- and school-level clustering (Online Appendix Table B.6), and both bootstrap schemes (Online Appendix Section B.6).

Figure 3: Decomposition of the reweighted common-support *Enseña Chile*–traditional value-added gap.



Notes: Additive components of the reweighted teacher-level gap (sum within rounding). *Enseña Chile* mean unweighted; sampled traditional teachers reweighted to the 2011 teacher-census PSU/PAA distribution over common-support deciles (Online Appendix B.8). The route block's dashed outline and labelled interval are the traditional-side stratified-bootstrap 95% CI on the route residual ($N = 104$).

teacher cohort, *Enseña Chile* recruits sit near the 97th percentile of ability (+1.9 SD above the mean teacher). Their predicted VA, pooled across first- and second-year teachers to avoid relying on small year-specific cells, falls at the 42nd percentile. Both figures are positions in the distribution of model-predicted VA, which carries no residual teacher-effect dispersion; they therefore rank teachers by predicted effectiveness and are not percentiles of realized national VA. That places them around the national median and close to the typical teacher at schools of comparable SES composition (51st percentile). In short, selection places *Enseña Chile* teachers near the top of the ability distribution, while their early-career status brings predicted effectiveness back toward the middle.

Student perception surveys are consistent with this pattern, though they provide only weak corroborating evidence. *Enseña Chile* classrooms score 0.17 points higher on the 5-point survey average, a statistically significant gap, but the classroom-level survey–VA correlation is small.[9]

4. Discussion and conclusion

The main empirical result is a decomposition of the reweighted common-support effectiveness gap between alternative-pathway and traditional teachers in Chile. In our linked data, that gap reflects two offsetting forces: a positive contribution from *Enseña Chile* teachers' pre-college cognitive ability ($+0.25\sigma$) and a first-year inexperience penalty (-0.30σ). What remains is a small route-of-preparation residual (-0.10σ), statistically indistinguishable from zero.

The residual is estimated from the early-career overlap revealed in the corrected experience data (Section 2) and from a modest sample. It is a remainder rather than a structural route parameter, absorbing route of preparation together with any ability-by-experience interaction the additive specification omits and any teacher-level unobservables (motivation, unmeasured selection) correlated with route. The decomposition is descriptive, not causal. The estimates are consistent with selection plus rapid early-career learning explaining most of the gap, in line with U.S. teacher preparation program (TPP) evidence that within-route variance exceeds mean differences between routes (von Hippel et al., 2016; von Hippel and Bellows, 2018) and with cross-national evidence that teacher cognitive skills predict student out-

[9]For more detail, see Online Appendix B.3.

comes (Hanushek et al., 2019). The result is best read as a failure to detect a route penalty rather than as definitive evidence of parity.

These findings speak to the margin between preparation routes, not to the value of preparation itself. They are consistent with the marginal year of formal pedagogical coursework adding little measured VA beyond pre-college ability and structured early-career learning. Nor is this a contrast between trained and untrained teachers. *Enseña Chile* provides four weeks of intensive pre-service instruction followed by two years of in-service coaching, peer learning, and tutor visits. The relevant comparison is therefore between a traditional model that front-loads preparation over four to five years and an alternative model that compresses pre-service training and combines it with on-the-job support. Evidence from China, where specialized teachers' colleges modestly outperform comprehensive-university preparation (Zheng et al., 2024), also suggests that practice-based pedagogy may add value. Our design cannot distinguish that content margin from total time in coursework.

For policy, the relevant counterfactual is a staffing decision: how an *Enseña Chile* teacher compares with the *marginal* hire who would otherwise take the slot, not the average traditional teacher or the unusually high-ability ones in our purposive sample. We do not observe that marginal hire, so what follows is an interpretation of the evidence, not a direct estimate.

Two facts make the staffing case plausible. First, *Enseña Chile* recruits sit near the 97th percentile of the national PSU-era teacher ability distribution, well above the teachers who typically staff comparable low-SES schools (Online Appendix Figure B.2)—and local labor markets are thinnest precisely where strong candidates are scarce (Meckes and Bascopé, 2012; Cabezas et al., 2017).

Second, Model 1 (no controls; Table 2) places *Enseña Chile* teachers at rough parity with the traditional teachers already working in these schools; setting first-year teachers aside moves the point estimate modestly in *Enseña Chile*'s favor (+0.35, s.e. 0.23, versus +0.28 for traditional teachers in the same schools), though neither contrast is statistically significant and the ordering does not survive the faculty-composition control.[10] Together with the $+0.41\sigma$-per-SD ability gradient, this *suggests* (though absent the counterfactual hire cannot establish) that placing an *Enseña Chile* teacher need not lower, and may raise, instructional quality where recruitment is hardest. The argument weakens wherever a high-PSU/PAA traditional alternative is readily available.

The SES breakdown is consistent with this interpretation, though the cells are small (Online Appendix Table B.3). *Enseña Chile* teachers are at rough parity with traditional teachers in the lowest-SES tier (Traditional VA −0.08 vs. *Enseña Chile* −0.11). In the low-mid and mid tiers, where the traditional comparison teachers are stronger in the raw data, *Enseña Chile* teachers have lower VA (−0.28 vs. +0.32 in low-mid; −0.34 vs. +0.06 in mid). The small *Enseña Chile* samples in those tiers (14 and 3 teachers, respectively) limit inference, but the pattern is consistent with selective placement of *Enseña Chile* in the most disadvantaged schools, where the counterfactual hire is most likely to have low PSU/PAA.

Several limitations should guide interpretation. First, the VA estimates are cross-sectional teacher fixed effects from a single school year, producing wide standard errors on 114 teachers. Second, the purposive sample describes this

[10]See Online Appendix Table B.1.

analytic group; voluntary school participation could still generate selection on unobservables, though recorded non-participation reasons were administrative (Online Appendix A) and the direction is ambiguous. Third, the first- to second-year *Enseña Chile* improvement spans two small cohorts (20 and 17 teachers) and cannot fully separate within-teacher learning from cohort selection, differential attrition, or transitory VA noise; a multi-year *Enseña Chile* cohort panel would address this.

Three implications follow for *Enseña Chile*. First, we fail to detect a route penalty: conditional on ability and experience, *Enseña Chile*'s compressed pre-service training plus in-service coaching yields VA statistically indistinguishable from the traditional four-to-five-year degree, though the data cannot rule out a moderate shortfall. Second, in the schools where *Enseña Chile* places teachers, its recruits can plausibly sustain, and may raise, instructional quality. Third, experience is the binding constraint. The program's two-year term means its teachers mechanically leave just as early-career returns are largest; retention beyond that commitment is therefore the margin with the most measured upside.

The boundary is equally important. This paper evaluates the *teachers Enseña Chile* supplies, not the *program*. We do not observe its effect relative to the teachers who would otherwise have been hired, its impact on the broader teacher pipeline, or participants' later trajectories in and out of education. Those are program-evaluation questions that our purposive, cross-sectional VA design is not built to answer.

The broader lesson is necessarily cautious, but it extends beyond *Enseña Chile*. For systems facing shortages in secondary mathematics, selective alternative pathways coupled with structured coaching may expand the supply of effective teachers without a detectable loss of instructional quality, provided they also confront the retention problem that fixed-term models create. On this evidence, the relevant policy margin is not simply whether four-year teacher colleges or short alternative routes produce more effective teachers on average. It is whether a system can attract high-ability candidates, support them through the steep early-career learning curve, and keep them in the classrooms where they are needed most.

CRediT authorship contribution statement

Conceptualization: SDG, CAN, SG. Methodology: SDG, SG. Formal analysis: SDG. Investigation: SDG. Data curation: SDG, CAN. Writing–original draft: SDG. Writing–review & editing: SDG, CAN, SG. Visualization: SDG. Supervision: CAN. Project administration: SDG. Funding acquisition: CAN.

Declaration of competing interest

The authors declare no competing interests.

Data availability

The aggregate data underlying every table and figure in the paper—group means, regression coefficients, decomposition components, binned and fitted values, and the national benchmark distribution—are compiled, together with the full analysis code, as a replication package available from the authors on request. The underlying individual-level records (student SEPA assessments, teacher PSU/PAA scores, and the MINEDUC *Docentes* experience panel) contain sensitive personal information and cannot be redistributed; they are available from the authors subject to data-sharing agreements with MIDE-UC (SEPA), MINEDUC/DEMRE (PSU/PAA), and Enseña Chile (program rosters).

Funding

This research was supported by a grant from the World Bank, administered through Teach For All. The World Bank and Teach For All supported the testing campaign and provided technical and logistical feedback; final analysis, interpretation, writing, and the decision to submit were the authors' responsibility.

Ethics approval

The 2016 data collection and the linked-records analysis reported here were both reviewed and approved by the Institutional Review Board of the University of Southern California, under a protocol held by S. De Gregorio. Data collection proceeded with the consent of participating schools and parents. The linked analysis uses de-identified assessment, survey, and administrative records governed by data-sharing agreements; reported results are aggregate and do not identify individual students, teachers, or schools.

Acknowledgments

We thank Enseña Chile for facilitating access to its network of teachers and schools, MIDE-UC for the SEPA assessments and external proctoring, Teach For All for administering the project grant, and the World Bank for financial support and technical feedback. We are grateful to Tomás Recart, Tomás Vergara, and Francisco Contreras (Enseña Chile); Laura Lewis, Robbie Dean, Evan Stockton, Kendra Hennig, Sarabeth Berman, and Anna Molero (Teach For All); Alonso Sánchez (World Bank); and Cecilia Rouse, Christina Gage, and Lisa Markman (Princeton University) for their support, and to Fabiola Alba Vivar, María Fernanda Ramírez, and Alejandra Aponte for excellent research assistance. We also thank the participating schools, teachers, and students. Any errors are our own.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used OpenAI's ChatGPT/Codex and Anthropic's Claude (Claude Code) to assist with drafting and language editing, formatting, and consistency and error-checking across the manuscript, its exhibits, and the replication code. The authors specified all analyses, verified all reported results against the underlying data, reviewed and edited the content, and take full responsibility for the content of the publication.

References

Allen, R., Allnutt, J., 2017. The impact of Teach First on pupil attainment at age 16. British Educational Research Journal 43, 627–646. doi:10.1002/berj.3288.

Araujo, M.C., Carneiro, P., Cruz-Aguayo, Y., Schady, N., 2016. Teacher quality and learning outcomes in kindergarten. The Quarterly Journal of Economics 131, 1415–1453.

Backes, B., Hansen, M., 2024. Estimates of Teach for America effects on student test and nontest academic outcomes over time. AERA Open 10, 1–16. doi:10.1177/23328584241234874.

Bietenbeck, J., Piopiunik, M., Wiederhold, S., 2018. Africa's skill tragedy: Does teachers' lack of knowledge lead to low student performance? Journal of Human Resources 53, 553–578.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., Wyckoff, J., 2008. The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. Journal of Policy Analysis and Management 27, 793–818.

Cabezas, V., Paredes, R., Bogolasky, F., Rivero, R., Zarhi, M., 2017. First job and the unequal distribution of primary school teachers: Evidence for the case of Chile. Teaching and Teacher Education 64, 66–78.

Chetty, R., Friedman, J.N., Rockoff, J.E., 2014. Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. American Economic Review 104, 2593–2632. doi:10.1257/aer.104.9.2593.

Chiang, H., Clark, M.A., McConnell, S., 2017. Supplying disadvantaged schools with effective teachers: Experimental evidence on secondary math teachers from Teach For America. Journal of Policy Analysis and Management 36, 97–125.

Citkowicz, M., Chen, C., Castro, M., Arellano, B., 2024. A Meta-analysis of Teach For America Teacher Impacts. Technical Report. American Institutes for Research. Arlington, VA.

Clark, M.A., Isenberg, E., 2020. Do Teach For America corps members still improve student achievement? evidence from a randomized controlled trial of Teach For America's scale-up effort. Education Finance and Policy 15, 736–760. doi:10.1162/edfp_a_00311.

Clark, M.A., Isenberg, E., Liu, A.Y., Makowsky, L., Zukiewicz, M., 2017. Impacts of the Teach For America Investing in Innovation Scale-Up. Technical Report. Mathematica Policy Research. Princeton, NJ.

Clotfelter, C.T., Ladd, H.F., Vigdor, J.L., 2007. Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. Economics of Education Review 26, 673–682.

Glazerman, S., Mayer, D., Decker, P., 2006. Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. Journal of Policy Analysis and Management 25, 75–96.

Hanushek, E.A., Piopiunik, M., Wiederhold, S., 2019. The value of smarter teachers: International evidence on teacher cognitive skills and student performance. Journal of Human Resources 54, 857–899.

Hanushek, E.A., Rivkin, S.G., 2010. Generalizations about using value-added measures of teacher quality. American Economic Review 100, 267–271.

Hill, H.C., Rowan, B., Ball, D.L., 2005. Effects of teachers' mathematical knowledge for teaching on student achievement. American Educational Research Journal 42, 371–406.

von Hippel, P.T., Bellows, L., 2018. How much does teacher quality vary across teacher preparation programs? reanalyses from six states. Economics of Education Review 64, 298–312. doi:10.1016/j.econedurev.2018.01.005.

von Hippel, P.T., Bellows, L., Osborne, C., Lincove, J.A., Mills, N., 2016. Teacher quality differences between teacher preparation programs: How big? how reliable? which programs are different? Economics of Education Review 53, 31–45. doi:10.1016/j.econedurev.2016.05.002.

Jackson, C.K., Rockoff, J.E., Staiger, D.O., 2014. Teacher effects and teacher-related policies. Annual Review of Economics 6, 801–825.

Jacob, B.A., Rockoff, J.E., Taylor, E.S., Lindy, B., Rosen, R., 2018. Teacher applicant hiring and teacher performance: Evidence from DC public schools. Journal of Public Economics 166, 81–97.

Ladd, H.F., Sorensen, L.C., 2017. Returns to teacher experience: Student achievement and motivation in middle school. Education Finance and Policy 12, 241–279.

Lavado, P., Guzmán, E., 2020. Evaluación de impacto de Enseña Perú: resultados en logros de aprendizaje (2013–2019). Technical Report. Universidad del Pacífico. Lima, Perú. Difference-in-differences with student test data.

Meckes, L., Bascopé, M., 2012. Distribución inequitativa de los nuevos profesores mejor preparados: características del empleador y nivel socioeconómico de la escuela como factores críticos. Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana 49, 1–19.

Metzler, J., Woessmann, L., 2012. The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation. Journal of Development Economics 99, 486–496.

Neilson, C., Gallegos, S., Calle, F., Karnani, M., 2023. Screening and recruiting talent at teacher colleges using pre-college academic achievement. Working paper, Yale University, Universidad Adolfo Ibáñez, University of Chicago Booth, and MIT.

Papay, J.P., Kraft, M.A., 2015. Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. Journal of Public Economics 130, 105–119. doi:10.1016/j.jpubeco.2015.02.008.

Rockoff, J.E., Jacob, B.A., Kane, T.J., Staiger, D.O., 2011. Can you recognize an effective teacher when you recruit one? Education Finance and Policy 6, 43–74.

Xu, Z., Hannaway, J., Taylor, C., 2011. Making a difference? The effects of Teach For America in high school. Journal of Policy Analysis and Management 30, 447–469.

Zheng, Q., Xie, X., Ye, X., Wei, Y., 2024. Do teachers colleges prepare more effective teachers? Evidence from a top school district in China. China Economic Review 87, 102225. doi:10.1016/j.chieco.2024.102225.