Teacher Preparation, Pre-College Human Capital and Student Learning

Evidence from *Enseña Chile*

Soledad De Gregorio Christopher A. Neilson Sebastián Gallegos

Teaching slides · 2026

Figures and numbers correspond to the August 2026 working paper.

The puzzle these programs pose

- Teach For America and the **Teach For All** network recruit high-achieving graduates, train them for a few weeks, and place them in hard-to-staff schools.
- Evaluations repeatedly find they do **about as well as** traditionally prepared teachers — sometimes modestly better in mathematics.
- That result is puzzling. Traditional preparation takes **four to five years**; the alternative route compresses pre-service training into weeks.

The question we can now ask

Is the parity explained by **who these programs select**, or by **what recruits learn on the job**? The two are almost always tangled together.

Teaching note: Good place to poll the room before showing any results. Students usually split, and the split itself motivates the decomposition.

Why the question stayed open

To separate selection from experience you need, for the *same* teachers:

1. a measure of **academic ability taken before any teacher training**;
2. **student test-score panel data** good enough to estimate value-added; and
3. **years of experience** measured accurately.

Almost no setting has all three. US studies use post-training subject knowledge, licensure tests, or pre-hire screening batteries — all administered to an *already-screened* pool.

Chile has all three, for a Teach For All affiliate.

Setting: *Enseña Chile* and the PSU

Enseña Chile (eCh)

- Chilean Teach For All affiliate.
- Four weeks of intensive pre-service training.
- Then **two years** of in-service coaching, peer learning, tutor visits.
- Recruits from selective universities.

The PSU/PAA entrance exam

- National, high-stakes university admission test.
- Taken **before** any teacher preparation — a clean pre-program ability measure.
- Mean 500, SD 110 among all test-takers.
- Among *teachers* the combined math–language SD is about **80 points**.

Teaching note: The 80-point teacher SD matters for every effect size below. Units are the single easiest thing to get wrong here.

Data: a testing campaign, then two administrative links

We ran the tests ourselves

SEPA mathematics assessment (MIDE-UC), administered by external proctors at the **start (May)** and **end (November)** of the 2016 school year, grades 9–11.

Students		3,756
Classrooms		169
Mathematics teachers	114	(37 eCh, 77 traditional)
Schools		59

Linked to **DEMRE** PSU/PAA scores and the Ministry of Education *Docentes* panel for years of service.

Teaching note: Worth dwelling on: value-added estimates are rare in middle-income countries precisely because this repeated classroom-level testing usually does not exist.

The sample is purposive — and that is deliberate

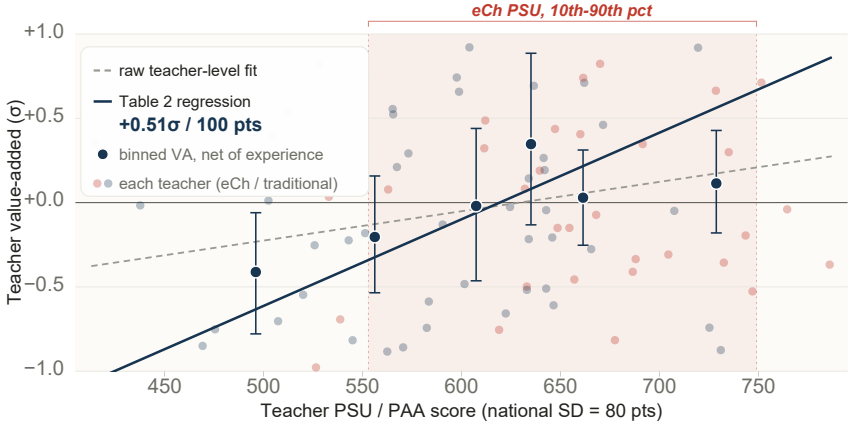
- eCh recruits cluster at the **top** of the PSU/PAA distribution.
- A nationally representative comparison group would therefore **confound route with ability** — the very thing we want to separate.
- So we stratified non-partner schools by their teachers' PSU/PAA decile and invited **at least one school per decile**.

The trade-off, stated plainly

We buy the **ability overlap** needed to estimate the gradient, and give up **representativeness**. Group means are analytic-sample comparisons, not population estimates for Chilean teachers.

Teaching note: A useful design discussion: when is a non-representative sample the *right* choice? Contrast with what a random sample would and would not identify.

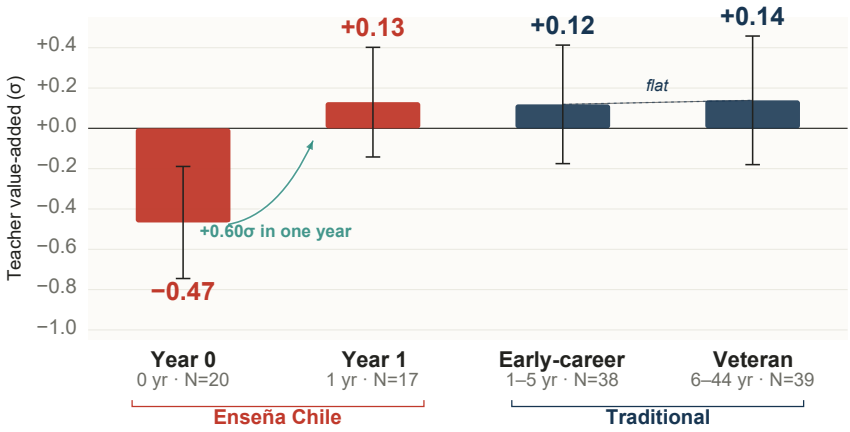
Result 1: pre-college ability predicts value-added



A 100-point higher PSU/PAA score predicts $+0.51\sigma$ higher value-added (classroom-clustered s.e. 0.12) — equivalently $+0.41\sigma$ per standard deviation of teacher ability, since one SD is about 80 points.

Teaching note: The units trap: 0.51 is *per 100 points*, not per SD. Students who quote 0.51 as a per-SD effect overstate it by about 25%.

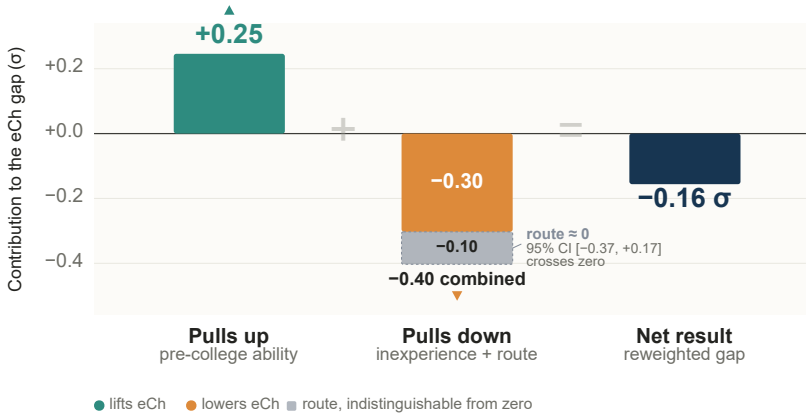
Result 2: the experience profile is steep, then flat



First-year eCh teachers average -0.47σ; second-year eCh teachers average $+0.13\sigma$ — the same as traditional teachers, whose own profile is flat from early career to veteran.

Teaching note: Press on this: the $+0.60\sigma$ difference compares *two different cohorts* measured in one testing

Result 3: decomposing the gap



Reweighting the sampled traditional teachers to the national common-support PSU/PAA distribution gives a gap of -0.16σ, which splits into $+0.25\sigma$ from pre-college ability, -0.30σ from first-year inexperience, and -0.10σ of route residual (95% CI $[-0.37, +0.17]$ — it crosses zero).

Reading the decomposition carefully

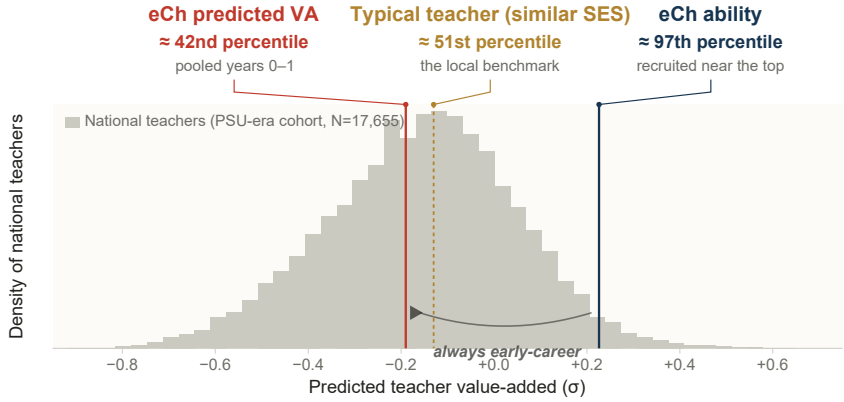
- The **net gap** is itself not statistically distinguishable from zero. What the exercise identifies is its **composition**.
- **Ability** and **inexperience** are each individually well determined; they pull in opposite directions and largely cancel.
- The **route residual is a remainder**, not a structural parameter. It absorbs training, motivation, unobserved selection, and any ability-by-experience interaction the additive form omits.

The honest summary

This is a **failure to detect** a route penalty — not evidence that the two training routes are equivalent. The data cannot rule out a moderate shortfall.

Teaching note: This slide is the one to keep if you cut the deck for time. The distinction between "no evidence of a difference" and "evidence of no difference" is the transferable lesson.

Recruited at the top, delivering at the middle



Against the PSU-era national teacher cohort ($N \approx 17{,}700$), eCh recruits sit near the **97th percentile** of ability, while their predicted value-added — pooled across their first two years — falls at the **42nd percentile**. Both are positions in the *model-predicted* distribution, not ranks in realized national value-added.

What this does and does not say about policy

Supported

- Selection on pre-college ability is a real, measurable margin.
- Where strong candidates are scarce, placing an eCh teacher need not lower instructional quality.
- Experience is the binding constraint: the two-year term ends just as early-career returns are largest.

Not supported

- That the two routes are equivalent.
- That the program *as a whole* is effective — we evaluate the teachers it supplies, not the program.
- Any comparison with the **marginal hire** a vacancy would otherwise draw. We never observe that teacher.

Teaching note: The marginal-hire point is the crux of most policy debates about these programs, and it is exactly what the design cannot deliver.

Limitations worth teaching

1. **Single year of data.** Value-added is cross-sectional teacher fixed effects, reported unshrunk, so estimates carry classroom-level noise. One σ is one SD of the *estimated* teacher-effect distribution — noise-inflated, which makes the gradients conservative.
2. **Small cohorts.** The first-to-second-year improvement rests on 20 and 17 teachers observed in the same wave; learning, cohort composition, and attrition cannot be fully separated.
3. **Purposive sampling.** Interpretation rests on within-comparison validity, not external representativeness.
4. **Thin within-school cells.** About 1.9 sampled teachers per school, so school fixed effects are infeasible; a leave-out faculty-composition control is the placement adjustment the design supports.

Questions for discussion

1. The route residual's confidence interval is $[-0.37, +0.17]$. What sample size would you need to rule out a 0.10σ shortfall? Is that study feasible?
2. The comparison group was chosen to span the ability distribution rather than to be representative. What does that buy, and what does it cost?
3. If selection explains the ability contribution, what policy follows — recruit differently, train differently, or retain differently?
4. The program's two-year commitment ends exactly when returns to experience are steepest. Is that a design flaw, or the price of attracting these recruits?
5. What would change your mind about the near-zero route residual?

Adapting these slides

- Every block is marked with a `\section` comment in the source and can be dropped independently.
- **Short version (20 min):** title, the puzzle, the data slide, Results 1–3, and "Reading the decomposition carefully".
- **Methods version:** keep the purposive-sampling and limitations slides and add the paper's Section 2 for the value-added construction.
- Figures live in `../figures/`; the four used here are `fig_psu_va_gradient`, `fig_experience_by_route`, `fig3_decomposition`, `fig_national_percentile`.
- **Please keep the units convention:** PSU/PAA coefficients are per 100 points; divide by 1.25 for per-SD.

Reference

Paper

De Gregorio, S., C. A. Neilson and S. Gallegos (2026). "Teacher Preparation, Pre-College Human Capital and Student Learning: Evidence from *Enseña Chile*." Working paper.

Manuscript, online appendix and these slides:
`https://www.christopher-neilson.com/work/eCh_Teachers.html`

Individual-level records are governed by data-sharing agreements with MIDE-UC, MINEDUC/DEMRE and *Enseña Chile* and cannot be redistributed. The replication package, including the aggregate data behind every exhibit, is available from the authors on request.