Online Appendix

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

Soledad De Gregorio Christopher A. Neilson Sebastián Gallegos

A Sample construction and sampling design

A.1 Sampling frame and response

Of the 53 schools in Chile employing at least one *Enseña Chile* mathematics teacher in grades 9–11 in 2016, 38 schools (72%) agreed to participate. The 15 non-participating schools either declined or had already administered SEPA at the end of 2015 and could not re-test in 2016. The 24 non-partner schools in the Metropolitan Region comparison sample were selected from a frame stratified into deciles of school-mean teacher PSU/PAA; participating non-partner schools cover all ten deciles.

A.2 Sampling design and selection into participation

The study sample is *purposive*, not representative. The design was intended to span the teacher ability distribution, placing high- and low-PSU/PAA traditional teachers alongside the *Enseña Chile* group so that the ability–value-added relationship could be estimated (Section 2 of the main text). Interpretation therefore rests on within-comparison validity, supported by baseline-score controls and the school faculty-composition control reported in Table B.6, rather than on representativeness. Unweighted teacher means and route gaps in the paper should be read as analytic-sample comparisons. A population comparison would require leaving the *Enseña Chile* cohort mean unweighted while reweighting or post-stratifying the sampled comparison teachers to the relevant national teacher distribution over PSU/PAA, experience, geography, and school type.

Among the 53 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 as part of normal operations. For the remainder, the project's 2017 final report records only that the school declined or did not give a definitive response to the invitation, which is an outcome rather than a reason and does not bear on quality-related selection in either direction. Because non-participating schools did not enter the testing campaign, their student outcomes were not collected, so the study data do not support a formal balance test on achievement outcomes. We therefore cannot rule out selection on unobservables within the comparison. The direction of any such selection is ambiguous: improvement-oriented schools may select into the study, while less-disadvantaged schools that already meet their assessment needs may select out.

A.3 Geographic distribution of partner schools in the analytic sample

The 35 partner schools that contribute teachers to the analytic sample span 8 of Chile's 15 regions: 11 schools in the Metropolitan Region, 6 in Region V (Valparaíso), 1 in Region I (Tarapacá), and 17 across Regions VIII (Biobío), IX (Araucanía), X (Los Lagos), XI (Aysén), and XIV (Los Ríos). The non-partner comparison group is concentrated in the Metropolitan Region by design, where 40% of

Chilean students are enrolled. Regional and school fixed effects are not feasible given the cell sizes; the robustness specifications instead control for school faculty composition.

B Additional regression and heterogeneity results

B.1 Excluding first-year *Enseña Chile* teachers

Table B.1 reports the main regression specifications after restricting Columns 2–5 to second-year *Enseña Chile* teachers; Column 1 reports a Model 2 specification estimated only on *Enseña Chile* teachers. Excluding first-year *Enseña Chile* teachers raises the staffing-comparison coefficient on the *Enseña Chile*-in-partner-school indicator from 0.16 to 0.35 in the pooled specification (Column 2), although the estimate is not statistically significant. Adding the school faculty-composition control reduces that point estimate and does not produce a statistically significant route contrast (Columns 3 and 5). The conditional specifications continue to show a strong PSU/PAA gradient and a flat average experience profile outside the early *Enseña Chile* years.

Table B.1: Additional regression models (excluding first-year *Enseña Chile* teachers in Columns 2–5)

	(1) eCh M2	(2) M1 no FY	(3) M1+Fac no FY	(4) M2 no FY	(5) M2+Fac no FY
Traditional, partner sch.		0.277	0.217	0.451	0.365
		(0.239)	(0.236)	(0.241)	(0.224)
Enseña Chile, partner sch.		0.353	0.192	0.184	-0.067
		(0.226)	(0.229)	(0.340)	(0.347)
PAA/PSU (per 100 pts)	0.433**			0.575***	0.550***
	(0.140)			(0.163)	(0.139)
Experience	0.458**			0.035	-0.002
	(0.156)			(0.034)	(0.037)
Experience^2				-0.0005	0.0006
				(0.001)	(0.001)
Faculty ability			0.346***		0.381***
			(0.103)		(0.100)
Mid-Low SES	-0.362*	-0.374*	-0.218	-0.482*	-0.339
	(0.167)	(0.190)	(0.192)	(0.191)	(0.179)
Mid SES	0.246	-0.253	-0.424*	-0.537	-0.712**
	(0.315)	(0.263)	(0.210)	(0.306)	(0.258)
Mid-High SES		0.059	-0.121	-0.362	-0.546
		(0.329)	(0.325)	(0.300)	(0.298)
Class size	0.037**	0.036*	0.022	0.046***	0.034*
	(0.012)	(0.016)	(0.016)	(0.013)	(0.014)
Grade	0.154	0.334**	0.263*	0.347**	0.295**
	(0.109)	(0.128)	(0.125)	(0.112)	(0.106)
Constant	-5.388***	-4.189**	-3.083*	-8.152***	-6.969***
	(1.494)	(1.312)	(1.326)	(1.291)	(1.182)
R^2	0.350	0.107	0.185	0.277	0.360
Observations	1,155	2,828	2,803	2,828	2,803
Faculty control	No	No	Yes	No	Yes

Notes: Column (1) is *Enseña Chile* teachers only; columns (2)–(5) drop first-year *Enseña Chile* teachers (no FY). Dependent variable is standardized value-added; 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 the coefficient is 0.8 times its value per SD. SES tiers are the SIMCE groups, with Low omitted. Experience is Docentes-panel corrected; SEs are clustered at the classroom. Faculty ability is leave-out school-faculty PAA/PSU percentile, standardized. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.2 reports a more direct teacher-level check of the route residual. It drops first-year *Enseña Chile* teachers entirely, so the *Enseña Chile* indicator compares second-year *Enseña Chile*

teachers with traditional teachers rather than using a first-year dummy to allocate the year-zero penalty. In this common-support comparison, the second-year *Enseña Chile* route residual is close to zero after conditioning on PSU/PAA and experience (-0.10σ with linear experience; -0.03σ after also adding school faculty ability), and is never statistically distinguishable from zero. The main near-zero route residual is therefore not masking a detectable second-year *Enseña Chile* shortfall.

Table B.2: Second-year route-residual checks after dropping first-year *Enseña Chile* teachers

	Raw	**+PSU**	**+PSU+Exp**	**+PSU+Exp^2**	**+PSU+Exp+Fac**
Second-year *Enseña Chile* route residual	0.065	-0.194	-0.105	-0.099	-0.031
	(0.182)	(0.232)	(0.265)	(0.281)	(0.254)
Observations	84	84	84	84	82

Notes: Teacher-equal OLS route-residual checks after dropping first-year *Enseña Chile* teachers; the reported coefficient is the *Enseña Chile* dummy, and robust HC1 SEs are in parentheses. Dependent variable is standardized value-added; columns add controls cumulatively (PSU/PAA, experience, experience^2, leave-out school-faculty ability). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

B.2 Value-added by school SES and teacher type

Table B.3 reports unweighted mean teacher value-added by school socioeconomic status and teacher type within the analytic sample. *Enseña Chile* teachers achieve parity with sampled traditional teachers in the lowest-SES tier (Low SES: traditional -0.08 vs. *Enseña Chile* -0.11), where local teacher labor markets are thinnest. In the low-mid and mid SES tiers, *Enseña Chile* teachers exhibit lower mean value-added than sampled traditional teachers, but the *Enseña Chile* samples are small (14 and 3 teachers respectively). *Enseña Chile* teachers do not operate in the mid-high SES tier, which is dominated by traditional teachers in non-partner schools; no high-SES school appears in the sample.

Table B.3: Teacher VA by School SES and Teacher Type

SES	**Traditional**	***Enseña Chile***	**Total**
Low	-0.08	-0.11	-0.10
N Teachers	22	20	42
N Students	571	567	1,138
Low-Mid	0.32	-0.28	0.12
N Teachers	29	14	43
N Students	1,013	530	1,543
Mid	0.06	-0.34	-0.02
N Teachers	12	3	15
N Students	400	58	458
Mid-High	0.13	.	0.13
N Teachers	14	0	14
N Students	617	0	617
Total	0.13	-0.19	0.02
N Teachers	77	37	114
N Students	2,601	1,155	3,756

Notes: Unweighted teacher-level mean value-added (standardized) by SIMCE socioeconomic group (`cod_grupo`) and route, within the analytic sample; eCh pools both *Enseña Chile* cohorts. Experience-corrected VA, consistent with Table 2 and the decomposition.

B.3 Student survey averages and correlations

Table B.4 reports student perception survey averages by teacher group across the eight MET-aligned dimensions on a 1–5 scale. *Enseña Chile* classrooms average 0.17 points higher than pooled traditional classrooms on the 1–5 survey average (final row of Table B.4), a statistically significant difference that is consistent with the VA results but offers only a weak independent test.

Table B.4: Student survey averages by teacher group.

Category	**Trad. non-partner**	**Trad. eCh-partner**	***Enseña Chile***	**Total**
1 – Student input and ideas valued	3.7 (0.93)	3.5 (1.00)	3.9 (0.85)	3.7 (0.92)
2 – Rigorous expectations held	3.8 (0.78)	3.6 (0.88)	3.9 (0.68)	3.8 (0.78)
3 – Positive culture and learning environment	3.2 (0.72)	3.1 (0.65)	3.1 (0.67)	3.1 (0.69)
4 – Encouraging and supportive relationships	3.4 (0.98)	3.4 (0.99)	3.6 (0.90)	3.5 (0.96)
5 – Understanding checked for and ensured	3.7 (0.81)	3.5 (0.86)	3.8 (0.71)	3.7 (0.79)
6 – Engaging learning environment	3.4 (1.03)	3.3 (1.03)	3.6 (0.95)	3.4 (1.01)
7 – Learning fully internalized	3.7 (0.91)	3.5 (1.00)	3.9 (0.79)	3.7 (0.90)
8 – Classroom participation	3.4 (0.84)	3.3 (0.91)	3.7 (0.77)	3.5 (0.84)
Survey Average	3.38 (0.87)	3.23 (0.95)	3.51 (0.88)	3.39 (0.89)

Notes: Student-level means on the 1–5 survey scale; SDs in parentheses. Dimensions are the eight Tripod/MET domains (negatively worded items reverse-coded).

Table B.5 reports the classroom-level correlation matrix between VA, the eight survey dimensions, and the survey total. It is computed over the 149 classrooms rather than over student records: VA and all eight survey dimensions are constant within a classroom, so correlating across students would replicate each classroom about twenty-two times and understate the standard errors. On that basis

only one VA correlation is distinguishable from zero, "positive culture and learning environment" ($r = 0.24$, $p = 0.003$); "rigorous expectations" is the next largest ($r = 0.15$) but is not significant at conventional levels ($p = 0.07$). The ordering is consistent with the Measures of Effective Teaching project's finding that these dimensions are among the most predictive of student achievement (Kane and Staiger, 2012, *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*), but with 149 classrooms the survey–VA evidence is weak on its own. The pattern is suggestive rather than conclusive: the *Enseña Chile* perception advantage in Table B.4 is largest on participation, engagement, and student input, but disappears on "positive culture and learning environment," where *Enseña Chile* classrooms score 3.1 and sit at or just below traditional classrooms. Because that dimension is also the most VA-predictive and most closely related to classroom-management experience, the survey evidence is consistent with the first-year inexperience penalty documented in the main text.

Table B.5: Correlations: Teacher VA and Student Survey Dimensions

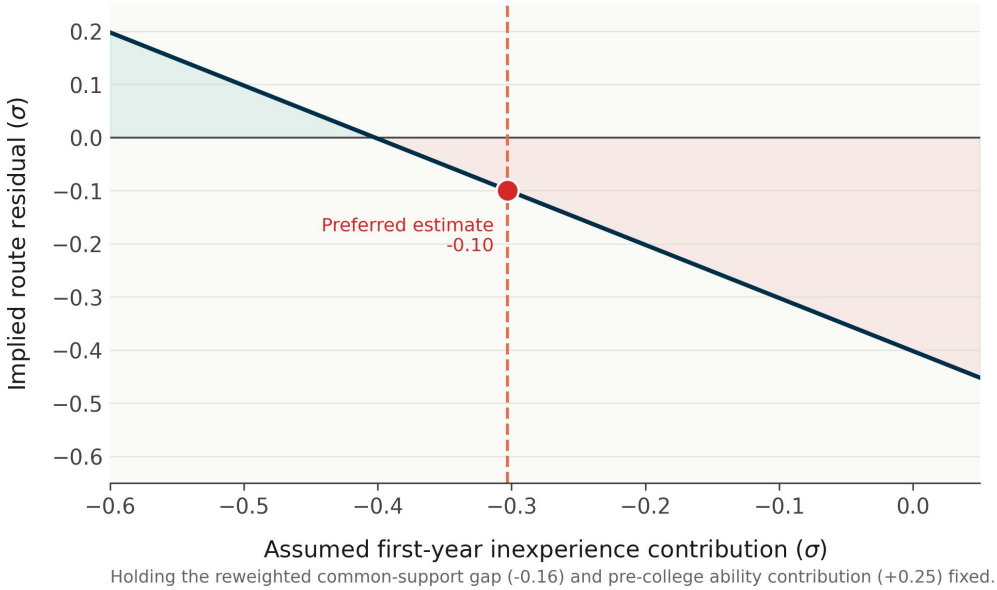
	VA	**Cat1**	**Cat2**	**Cat3**	**Cat4**	**Cat5**	**Cat6**	**Cat7**	**Cat8**	**Total**
Teacher Value-Added	1									
1- Student input (Cat1)	0.035	1								
2- Rigorous expectations (Cat2)	0.15	0.93*	1							
3- Positive culture (Cat3)	0.24*	0.50*	0.54*	1						
4- Supportive relationships (Cat4)	-0.037	0.92*	0.87*	0.41*	1					
5- Understanding ensured (Cat5)	0.049	0.91*	0.91*	0.56*	0.88*	1				
6- Engaging environment (Cat6)	0.064	0.91*	0.88*	0.54*	0.88*	0.94*	1			
7- Learning internalized (Cat7)	0.021	0.92*	0.93*	0.46*	0.91*	0.92*	0.91*	1		
8- Classroom participation (Cat8)	-0.024	0.87*	0.86*	0.41*	0.90*	0.84*	0.84*	0.92*	1	
Total	0.068	0.83*	0.84*	0.51*	0.78*	0.81*	0.81*	0.83*	0.76*	1

Note: $N = 149$. * significant at $p < 0.05$.

B.4 Sensitivity of the route residual to the inexperience penalty

Figure B.1 holds fixed the reweighted common-support gap (-0.16σ) and the pre-college ability contribution ($+0.25\sigma$), then varies the assumed first-year inexperience contribution. The preferred estimate uses the weighted decomposition's first-year inexperience contribution (-0.30σ), which implies a route residual of -0.10σ. The figure shows that the residual becomes more negative if the true first-year penalty is smaller in magnitude, and becomes positive only if the first-year penalty is substantially larger than the preferred estimate.

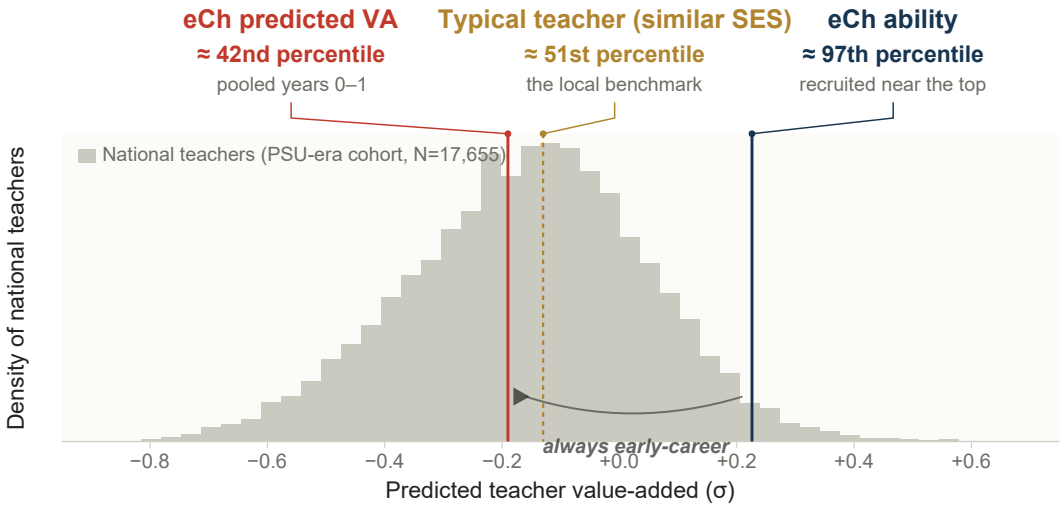
Figure B.1: Sensitivity of the route residual to the assumed first-year penalty.



B.5 National teacher-distribution benchmark

Figure B.2 benchmarks *Enseña Chile* teachers against the national PSU-era teacher cohort. This placement is descriptive rather than a staffing counterfactual: it projects the within-sample ability–value-added model onto the national teacher distribution and compares the predicted ability position of the always-early-career *Enseña Chile* teachers with their predicted value-added once the first-year term is applied. Both sides of that comparison are fitted values. The reference distribution is the model's prediction for each national teacher, so it carries no residual teacher-effect dispersion and is narrower than the distribution of realized value-added; percentiles in it rank teachers by predicted effectiveness and should not be read as percentiles of realized national value-added.

Figure B.2: *Enseña Chile* teachers in the national teacher value-added distribution.



Notes: Predicted value-added is standardized to national-teacher SDs for the PSU-era teacher cohort, the 2011 teacher census linked to DEMRE entrance scores, $N \approx 17{,}700$. *Enseña Chile* teachers are recruited near the top on ability (97th percentile, +1.9 SD above the mean teacher), but pooled across their first two years their predicted value-added falls at the 42nd percentile, around the median and close to the typical teacher at schools of comparable SES composition (51st). Both percentiles are positions in the model-predicted distribution, not in the distribution of realized value-added.

B.6 Robustness summary

Across the robustness exercises, the central decomposition result is unchanged: (i) the unweighted raw eCh–traditional gap of -0.32σ is concentrated in first-year teachers; (ii) the PSU–VA gradient remains in the 0.43–0.58 range per 100 PSU/PAA points (0.34–0.46σ per SD of teacher ability) across specifications, including the *Enseña Chile*-only regression in Table B.1, column 1; and (iii) after reweighting sampled traditional teachers to the national common-support PSU/PAA distribution, the main gap is -0.16σ and the route residual remains statistically indistinguishable from zero. The main reweighted decomposition uses a 2,000-replicate traditional-side stratified bootstrap; the unweighted teacher-level checks use a teacher bootstrap. Because the traditional-side scheme holds the evaluated *Enseña Chile* cohort fixed, the component intervals reported in the main text are conditional on that cohort. Re-running the reweighted decomposition under an *Enseña Chile*-inclusive scheme—resampling the *Enseña Chile* cohort with replacement while keeping the traditional side stratified by common-support decile, which preserves the weight map in every replicate—widens the ability interval from $[+0.06, +0.42]$ to $[+0.02, +0.47]$ and the route interval from $[-0.37, +0.17]$ to $[-0.52, +0.28]$: neither conclusion changes, with the ability contribution remaining positive and the route residual continuing to cross zero. The first-year inexperience component is different in kind. Under the conditional scheme its interval is degenerate—an order of magnitude too narrow—because no sampled traditional teacher is in their first year, so the experience differential is the *Enseña Chile* first-year share itself, which that scheme holds fixed; this is why the main text reports no conditional interval for it. The *Enseña Chile*-inclusive scheme restores genuine sampling uncertainty for this component, giving $[-0.55, -0.09]$, which excludes zero; Section B.4 complements this by varying the assumed first-year penalty directly. These procedures propagate teacher-sampling and decomposition uncertainty but do *not* re-estimate the first-stage

value-added generation, so they omit generated-regressor uncertainty from estimated VA. Table 2 standard errors are clustered at the classroom; because route, PSU, and experience vary only at the teacher level, Table B.6 reports the headline coefficients re-clustered at the teacher and school levels.

B.7 Clustering, school sorting, and the faculty-composition control

Table B.6 reports Model 2 coefficients with standard errors clustered at the classroom (as in the main text), the teacher, and the school. Because route, PSU/PAA, and experience vary at the teacher level, teacher- and school-clustered SEs are the conservative benchmark. The lower panel adds the school faculty-composition control used in the main text: the leave-out mean national PSU/PAA percentile of a school's other teachers.

Moving from classroom- to teacher- and school-clustered SEs widens the route SEs by 20–35%: the marginally significant pooled +0.46 on the traditional-teacher-in-partner-school indicator (*no-yes*) at classroom clustering becomes statistically indistinguishable from zero once SEs reflect teacher- and school-level variation in route assignment. The PSU/PAA gradient remains highly significant at all three levels. Adding the school faculty-composition control leaves the gradient and route terms essentially unchanged (PSU/PAA +0.49 per 100 points; *Enseña Chile* in partner school −0.01). Faculty composition itself enters at +0.21 (classroom-clustered s.e. 0.09), significant at the 5% level, but it is not significant at conventional levels once the conservative teacher- and school-clustered SEs are used (0.12 and 0.13, giving $p \approx 0.09$ and 0.12). School fixed effects are not feasible at this sample size (58 schools, ≈ 1.8 sampled teachers each, 40% with one sampled teacher); the continuous faculty-composition control is therefore the more stable way to hold school-level sorting constant.

School fixed effects absorb the route levels, so the only route quantity they identify is the within-school *Enseña Chile*–traditional contrast. Estimated on the schools that sample both routes, that contrast is +0.061 (teacher-clustered s.e. 0.247): near zero, like the pooled route residual in the main text, and far too imprecise to distinguish from it. The same design limit applies to the PSU/PAA gradient. Splitting PSU/PAA into its school-mean and within-school components gives a large between-school coefficient (+0.732, teacher-clustered s.e. 0.221) and a smaller, statistically insignificant within-school one (+0.161, s.e. 0.133); the two are not distinguishable from each other at conventional levels. With roughly two sampled teachers per school and 40% of schools contributing a single teacher, the within-school comparison carries little information, so this split cannot separate individual ability from school-level sorting in either direction. We therefore report the continuous faculty-composition control as the placement adjustment the design supports, and read the ability contribution as an association within the analytic sample.

Sample for Table B.6: 3,454 students, 104 teachers, 156 classrooms, 58 schools.

Table B.6: Model 2 coefficients under alternative clustering.

	Classroom-clustered (main text) (156 clusters)	Teacher-clustered (104 clusters)	School-clustered (58 clusters)
PSU/PAA (per 100 pts)	$+0.514^{***}$	$+0.514^{**}$	$+0.514^{**}$
	(0.123)	(0.167)	(0.184)
eCh in partner school (*yes-yes*)	+0.116	+0.116	+0.116
	(0.288)	(0.375)	(0.386)
Traditional in partner school (*no-yes*)	+0.457	+0.457	+0.457
	(0.241)	(0.297)	(0.326)
Experience (years, *expv*)	+0.039	+0.039	+0.039
	(0.033)	(0.041)	(0.034)
Adding the school faculty-composition control (Model 2 + faculty; $N = 3{,}398$):			
School faculty ability (*facz*)	$+0.207^{*}$	+0.207	+0.207
	(0.086)	(0.122)	(0.133)
with PSU/PAA now $+0.49^{***}$, *yes-yes*	-0.006 (s.e. 0.31, classroom-clustered)		

Notes: Standard errors are clustered at the classroom (main text), teacher, and school levels (upper panel); upper-panel coefficients are identical across columns, and only the standard errors differ. The PSU/PAA coefficient is per 100 points of the entrance exam; one standard deviation of teacher PSU/PAA is about 80 points, so +0.514 per 100 points is $+0.41\sigma$ per SD. Model 2 controls also include *exp2*, SES tier dummies, class size, and grade. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. The lower panel adds the school faculty-composition control, the leave-out mean national PSU/PAA percentile of a school's other teachers; its coefficient is reported under the same three clustering levels, and the route and PSU/PAA terms are essentially unchanged by its inclusion.

B.8 Reweighting the sampled traditional teachers to the national PSU/PAA distribution

The main text reports unweighted analytic-sample means because the comparison sample was deliberately selected to span PSU/PAA ability rather than to represent the national teacher population. For a descriptive population comparison, the *Enseña Chile* mean remains unweighted: the 37 *Enseña Chile* teachers are the cohort evaluated here. Reweighting is applied only to sampled traditional teachers. Table B.7 reweights sampled traditional teachers to the national 2011 teacher-census PSU/PAA percentile distribution using common-support deciles. Because sampled traditional teachers do not cover the bottom two national PSU/PAA deciles, the table avoids extrapolation and reports the common-support comparison only.

Applying the same weights to the decomposition sample with observed raw PSU/PAA scores ($N = 104$) gives the main-text identity $-0.16\sigma = +0.25\sigma - 0.30\sigma - 0.10\sigma$ (components sum to the gap within rounding); one sampled traditional teacher in Table B.7 has a national percentile but lacks the raw score used in that decomposition.

Table B.7: Reweighting sampled traditional teachers to the national PSU/PAA distribution

	N	**Mean PSU/PAA pct.**	**Mean VA**	***Enseña Chile* - comparison**
Enseña Chile cohort, unweighted	37	91.1	-0.193	–
Sampled traditional, unweighted	68	77.7	+0.060	-0.252
Sampled traditional, reweighted	68	60.3	-0.038	-0.155
National target, common support	146,161	59.6	–	–

Notes: The *Enseña Chile* mean is not reweighted: it is the cohort mean for the evaluated *Enseña Chile* teachers with observed PSU/PAA. The sampled traditional teachers are reweighted to the national 2011 teacher-census PAA/PSU percentile distribution over common-support deciles (deciles 3–10); 11.8% of scored national teachers lie outside this support and are not extrapolated. The reweighted gap's traditional-side stratified-bootstrap 95% interval is [-0.361, +0.045].