From Classrooms to Adulthood: Long-Term Impacts of Teacher and Student Incentives in Mexico*

Jere R. Behrman† Susan W. Parker‡ Jorge Pérez Pérez§ Mauricio Romero¶
Petra E. Todd|| Kenneth I. Wolpin** Weilong Zhang††

March 26, 2026

Abstract

This paper evaluates the long-term effects of the Aligning Learning Incentives (ALI) experiment, which randomly assigned 88 federal high schools to a control group and three treatment arms — student incentives, teacher incentives, and combined incentives — with approximately 97,000 students participating across three academic years (2008–2011). Previous evaluations demonstrated short-run test-score gains of 0.2–0.3σ for treatment groups that offered student payments, but could not determine how much of the impact reflected increased test-taking effort rather than enhanced learning. We link the original experimental data with national administrative records on standardized tests, professional degree registries, social program enrollment, formal-sector employment, and voter rolls to trace effects over a 10–15 year horizon. Student incentives increased post-secondary degree attainment by approximately 4–5 percentage points (off a 22% baseline) and reduced enrollment in federal welfare programs by approximately 4–7 percentage points. These findings suggest that performance-based incentives during adolescence can generate positive lasting effects.

Keywords: performance incentives, teacher incentives, student incentives, long-term effects, human capital, randomized experiment, Mexico
JEL Codes: I21, I25, I28, J24

*Romero gratefully acknowledges financial support from the Asociación Mexicana de Cultura, A.C. and the Jacobs Foundation. All errors are our own.
†University of Pennsylvania; jbehrman@sas.upenn.edu
‡University of Maryland; swparker@umd.edu
§Banco de Mexico; jorgepp@banxico.org.mx
¶ITAM; J-PAL; mtromero@itam.mx
||University of Pennsylvania; ptodd@sas.upenn.edu
**University of Pennsylvania
††University of Cambridge; weilong.zhang@econ.cam.ac.uk

1 Introduction

Monetary incentives for teachers and students can raise test scores in the short run (e.g., Glewwe, Ilias, and Kremer (2010); Muralidharan and Sundararaman (2011); Barrera-Osorio and Raju (2017); Mbiti et al. (2019); Loyalka, Sylvia, Liu, Chu, and Shi (2019); Mbiti, Romero, and Schipper (2023); Fryer (2011)). However, evidence on whether these gains persist remains limited, with a few notable exceptions (see Lavy (2020); S. Cohodes, Eren, and Ozturk (2023)). Because incentives tied to exam performance may induce greater effort during test-taking rather than true learning, their effects on sustained human capital accumulation are uncertain.

This paper evaluates the medium- and long-run impacts of the Aligning Learning Incentives (ALI) experiment, a large-scale randomized incentive program that paid students, teachers, or both based on performance on externally administered mathematics tests in Mexican upper-secondary schools. The experiment randomly assigned 88 federal high schools, between 2008 and 2011, to one of four conditions: (T1) student incentives, (T2) teacher incentives, (T3) combined schoolwide incentives, or (C) control. Incentives were tied to proficiency categories on curriculum-aligned standardized exams, rewarding both achievement levels and year-to-year improvement. The combined arm further introduced group and peer components to promote coordination across teachers and students. The program reached over 100,000 students across five entry cohorts, making it one of the largest teacher-and-student incentive experiments at the secondary level in terms of student enrollment (Behrman, Parker, Todd, & Wolpin, 2015). Short-run impact estimates revealed the largest average effects for T3 (approximately 0.2–0.3σ), smaller effects for T1, and no impact for T2.

We link the original experimental data with five administrative sources to trace effects over a 10–15 year horizon: (i) the national standardized ENLACE Grade 12 exam, used as a low-stakes achievement measure at the end of high-school and proxy for high school graduation; (ii) the national registry of professional degrees (*cédulas profesionales*) through 2025; (iii) the 2019 federal social-program beneficiary database, covering *Prospera*, *Seguro Popular*, and *IMSS-Bienestar*; (iv) monthly formal-sector employment records from the Mexican Social Security Institute (IMSS); and (v) voter registry and turnout data from the National Electoral Institute (INE, 2015–2021). We estimate intent-to-treat effects for each treatment arm, allowing

for heterogeneous effects across cohorts and program exposure duration.

Our analysis yields four main results. First, student performance on the 12th-grade ENLACE standardized test — which was not tied to incentives — did not improve. Students in combined-incentive schools (T3) were ∼4 percentage points less likely to take the Grade 12 ENLACE exam (off a 70% base), and students receiving individual incentives (T1) show modestly lower mathematics scores conditional on test-taking. These patterns are consistent with evidence that incentive effects are strongest on directly incentivized exams and may not transfer to low-stakes testing environments (Levitt, List, Neckermann, & Sadoff, 2016; Gneezy et al., 2017; Mbiti et al., 2019, 2023). Second, students who received individual incentive payments were roughly 5 percentage points more likely to hold a post-secondary degree by 2025, relative to the 22% control mean. Third, student incentives reduced participation in federal social programs by approximately 7 percentage points from a 57% base, with the decline concentrated in *IMSS-Bienestar* and *Seguro Popular* — programs serving individuals without formal-sector employment. This is consistent with higher rates of formal labor market integration among treated students. Preliminary results from IMSS administrative records suggest positive effects on the likelihood of formal employment. Fourth, the combined student-and-teacher incentive arm reduced voter turnout by about 4 percentage points in the 2018 presidential election.

Our study is related to a small literature that examines the long-term effects of incentives on students' adult outcomes, primarily incentives given to teachers based on their students' performance. Muralidharan and Sundararaman (2011) present results from a five-year long randomized evaluation of group and individual teacher performance pay programs implemented in rural schools in one state in India. Students who completed five years of primary school during the time of the teacher incentive program performed significantly better on math and language tests, for which teachers received incentives, as well as in other subjects for which no incentives were given in comparison to students in control schools. The effect sizes were increasing in the duration of program exposure. Another finding was that the effect sizes associated with the group teacher incentives were smaller than those associated with the individual incentives.

Lavy (2020) analyzes the long-term effects of a teacher pay-for-performance program randomized in 40 Israeli high schools two decades prior. The analysis is based on an intervention

that randomized teacher bonuses for student test achievements. A decade after the program ended, treated students were 5 percentage points more likely to enroll in university and completed an additional 0.25 years of university education — a 30% increase relative to the control mean. Lavy (2020) also finds a 1.3 percentage-point increase in employment and an 8–9% increase in earnings at ages 28–30. No effects are found on marriage or fertility.

S. R. Cohodes, Eren, and Ozturk (2026) examine the Teacher Advancement Program (TAP) across approximately 100 schools in South Carolina, which combined incentive pay with teacher observation and professional development. Exploiting quasi-random variation in program adoption timing across South Carolina schools, the study finds that the program increased educational attainment, reduced criminal activity, and lowered dependence on government assistance in early adulthood. The estimated benefits exceed costs by a factor of three to five. More broadly, Chetty, Friedman, and Rockoff (2014) show that high-value-added teachers improve students' long-term outcomes, including college attendance and adult earnings, establishing that classroom-level variation in teaching quality can have lasting effects.

To our knowledge, no existing study compares the long-term effects of student, teacher, and combined incentive designs within a single factorial experiment over a horizon exceeding a decade. Our study extends existing evidence in three ways. First, we provide rare long-horizon evidence (10–15 years) on whether performance incentives during adolescence influence adult outcomes. Second, we compare the relative effectiveness of student- versus teacher-based incentives within a unified experimental framework involving approximately 97,000 students. Third, we demonstrate the value of linking experimental data to multiple administrative records — test scores, professional degrees, social programs, formal employment, and voter rolls — for tracing the life-cycle impact of educational interventions.

2 The Mexican education system

The Mexican education system operates under a federal model. The Ministry of Education (*Secretaría de Educación Pública*, SEP) sets policy and standards, while state governments are responsible for administration and day-to-day management. Education is structured into three levels: Basic, Upper Secondary, and Higher Education. Basic education, comprising preschool (3 years), primary (6 years, grades 1–6), and lower secondary (3 years, grades 7–9), is compulsory and

tuition-free. Upper secondary (grades 10–12) provides academic or vocational training. During the intervention period (2008–2011), upper secondary education was not compulsory, but access was open to all students who had completed lower secondary education.[1]

Spending on education in Mexico is comparable to that of other middle-income countries at 4.6% of GDP (World Bank, 2017a). Public schools account for nearly 90% of enrollment. Almost all students complete primary education (World Bank, 2017b), although learning outcomes are poor. For example, in 2018, fewer than half of students achieved basic proficiency in math or Spanish according to national assessments (Instituto Nacional para la Evaluación de la Educación, 2018). Disparities are especially acute in marginalized communities, where only one-third of students reach basic proficiency by the end of primary school. In the 2022 PISA test score rankings, Mexico ranked 51st out of 81 participating countries.

Over the last two decades, Mexico has implemented several reforms to improve access to education and enhance its quality. These include changes to teacher hiring, training, evaluation, and salaries. Teacher incentives — either based on students' performance on standardized assessments or tied to teachers' performance in pedagogical evaluations — are common in the Mexican context. Given the central role that teachers play in student learning, it is essential to assess the long-term consequences of these incentive schemes. This paper analyzes the medium- and longer-term effects of teacher incentives, student incentives, and their combination.

3 The Aligning Learning Incentives (ALI) experiment

The Aligning Learning Incentives (ALI) program was a three-year randomized controlled trial (2008–2009 to 2010–2011) involving 88 public high schools in Mexico. We randomly assigned schools to one of three incentive-based treatment groups or a control group. The school-level randomization was stratified by school size and prior graduation rates.

We assigned 20 schools to each of the three treatments: Treatment 1 (student incentives only), Treatment 2 (teacher incentives only), and Treatment 3 (combined incentives for students, teachers, and staff), with 28 schools designated as controls. Over 100,000 students across these schools participated in the experiment ($\sim$12,000 per treatment group and $\sim$17,000 in the control group

[1] A 2012 reform made upper secondary education compulsory nationwide (Santiago, McGregor, Nusche, Ravela, & Toledo, 2012), but enforcement continues to vary by state.

in the first year). In each of the three years, all students enrolled in grades 10, 11, and 12 in the treated schools were included in the program.

Below, we describe the intervention, sampling, randomization, and the student cohorts exposed to the experiment. Further information can be found in (Behrman et al., 2015).

3.1 The ALI Program and Sampling Scheme

The ALI intervention provided monetary rewards based on student outcomes on standardized, curriculum-based mathematics tests administered at the end of each school year (for grades 10, 11, and 12). Incentive payments were tied to students' test performance and to improvements in test scores, aligning the financial incentives of students and teachers with demonstrated learning gains.

Table 1 shows the incentive payments students could earn for each grade, depending on their initial performance category.[2] Student results on the ALI tests were classified into four performance categories: Pre-Basic, Basic, Proficient, and Advanced, which mirrors the categorization that the government uses in reporting the official ENLACE (Evaluación Nacional de Logro Académico en Centros Escolares) test score results; ENLACE was the national standardized test in use at the time. ENLACE tests are administered in 1st-9th grade and in 12th grade.

Incentives were paid based on end-of-year test outcomes relative to each grade's baseline proficiency. In the first year, the baseline proficiency was based on performance on the 9th-grade ENLACE test.

3.1.1 Treatment T1 - Student Incentives

Students received cash awards based on their own end-of-year proficiency, conditional on their start-of-year proficiency category (from the four categories above). Grade 10 and Grade 11 incentive schedules rewarded both category maintenance in the upper categories and upward movement; Grade 12 incentives rewarded only achieving test scores in the Proficient/Advanced categories, to align with graduation proficiency goals. Payments increased with greater improvements (reflecting higher effort costs from low baselines).

[2]The exams were designed by CENEVAL, a civil society organization similar to the Educational Testing Service in the United States. CENEVAL drew on the expertise of Mexican specialists in upper-secondary mathematics and closely followed the official curriculum for each grade.

Table 1: Student incentive schedule (pesos) by starting and ending proficiency

Start of grade	End of grade			
	Pre-Basic	**Basic**	**Proficient**	**Advanced**
Grade 10				
Start: Pre-Basic	0	4,000	9,000	15,000
Start: Basic	0	2,500	7,500	13,500
Start: Proficient	0	0	6,000	12,000
Start: Advanced	0	0	4,500	10,500
Grade 11				
Start: Pre-Basic	0	4,000	9,000	15,000
Start: Basic	0	0	7,500	13,500
Start: Proficient	0	0	6,000	12,000
Start: Advanced	0	0	4,500	10,500
Grade 12				
Start: Any	0	0	5,000	10,000

Notes: Amounts reflect the program's official payment matrix (Behrman et al., 2015).

3.1.2 Treatment T2 - Teacher Incentives

Mathematics teachers earned per-student bonuses equal to 5% of the student schedule, with a modest penalty when a student regressed and ended below Proficient/Advanced in Grades 10 and 11. A teacher's annual bonus was calculated across all tested students, with the total bounded below by zero. The teacher payment schedule, shown in Table 2, weighs gains by lower-baseline students more heavily to discourage cream-skimming.

Table 2: Teacher per-student schedule (pesos) by starting and ending proficiency

Start of grade	End of grade			
	Pre-Basic	**Basic**	**Proficient**	**Advanced**
Grade 10				
Start: Pre-Basic	0	200	450	750
Start: Basic	−125	125	375	675
Start: Proficient	−125	−125	300	600
Start: Advanced	−125	−125	225	525
Grade 11				
Start: Pre-Basic	0	200	450	750
Start: Basic	−125	0	375	675
Start: Proficient	−125	−125	300	600
Start: Advanced	−125	−125	225	525
Grade 12				
Start: Any	0	0	250	500

Notes: Amounts reflect the program's official payment matrix (Behrman et al., 2015). Totals are bounded below by zero.

3.1.3 Treatment T3 - Combined/Group Incentives

The third treatment offered multi-level incentives to encourage both individual effort and group cooperation. Students earned rewards based on their own test results and on their classmates' collective performance. Math teachers earned rewards based on both their students' outcomes and the overall performance of all math classes in the grade. Additionally, non-math teachers and school administrators in this treatment received bonuses tied to the school-wide student performance on the math exam. In essence, this treatment created a combination of individual and collective incentives, as everyone in the school had a stake in improving overall student math achievement.

More specifically, for students in grades 10 and 11, rewards combined an individual component identical to the Treatment 1 schedule with a group component equal to 1% of the total payments earned by classmates in the same mathematics class. In grade 12, only the individual component was applied to emphasize proficiency by the time of graduation. The design of this incentive schedule aimed to generate positive peer externalities.

Mathematics teachers were rewarded both for their students' outcomes, using the Treatment 2 formula, and for their peers' performance, receiving 25 percent of the average payment of other math teachers in the school. This second component was intended to encourage collaboration, such as mentoring younger colleagues and sharing teaching strategies.

Non-mathematics teachers also participated, each earning 25% of the schoolwide average payment for math teachers. Finally, principals and associate principals received 50% and 25%, respectively, of the same average. See Appendix B in (Behrman et al., 2015) for the detailed formulas used to calculate payments for students, math teachers, and other school staff. Together, these overlapping incentives created a system in which every adult in the school and every student in math classes had a direct stake in improved performance, fostering both individual motivation and collective responsibility.

3.2 Sample and randomization

We conducted the experiment in federal public high schools.[3] We applied several inclusion criteria before randomization. From the 706 federal high schools nationwide, we excluded schools with multiple daily sessions (morning/afternoon/evening shifts) and retained only

[3]There are also public schools managed by the states as well as private schools.

"morning-session" schools in which all students attend in a single shift.[4] We also excluded very new schools (in their first year of operation), schools with fewer than 200 or more than 2,000 students, schools with multiple campuses, and schools in the states of Oaxaca or Michoacán (due to logistical and data access constraints). To avoid spillovers, we excluded schools within a 10-mile radius of any other federal high school. As a result, the final sample is skewed toward more geographically isolated schools.

After applying these filters, 135 schools remained eligible for consideration. From these, 88 were randomly chosen for the experiment. We assigned schools to T1, T2, T3, or control via block randomization: schools were grouped into 9 blocks based on size and prior graduation rates, then within each block, schools were randomly assigned to one of the four groups.[5]

The experiment lasted three academic years (from 2008–09 to 2010–11). All students enrolled in these high schools (grades 10–12, typically ages 15–18) participated in the experiment. However, students were exposed to the programs for different durations, depending on their high school grade level when the program started or ended. Below, we describe five distinct cohorts that participated in the experiment.

3.3 Cohorts exposed to the experiment

Given the grade levels of students at the start of the ALI intervention (2008-09 academic year) and assuming normal grade progression (one grade per year, with grade 12 as the final year of high school), we can project when each cohort would be expected to graduate from high school. Students may not actually graduate on this projected date, though, due to grade failures or to dropping out.

12th Grade in 2008-09 (Cohort 1 or class of 2009): These students were in their final year of high school during the first year of the ALI program. They should have graduated in 2009 (at the end of the 2008–09 school year). Most of them (over 88%) were born in either 1990 or 1991. This cohort was exposed to the incentive program for only one year, as they completed high school at the end of the first treatment year. We refer to this cohort as the class of 2009.

[4]Multi-shift schools were dropped because each shift functions like a separate school with its own principal. However, teachers and students live in the same community, which could create issues if different shifts had different treatment status.

[5]Six randomizations were originally performed, and the "best" among the six in terms of covariate balance (based on a few covariates that were observable prior to data collection (e.g., school size, dropout rate)) was selected. The computer code with the exact assignment mechanism has since been lost, so re-randomization inference is not possible. In practice, this implies that conventional OLS standard errors are likely conservative, as trimming the tails of the treatment/control distribution reduces variance across randomizations.

11th Grade in 2008-09 (Cohort 2 or class of 2010): These students had two years of high school remaining at the start of the program. They should complete 12th grade in the 2009–10 academic year and graduate in 2010. Most of them (over 87%) were born in either 1991 or 1992. This group received two years of exposure to the ALI incentives — in 11th grade and 12th grade — before completing high school. We refer to this cohort as the class of 2010.

10th Grade in 2008-09 (Cohort 3 or class of 2011): These students were in the first year of high school when the program began. They are expected to graduate in 2011, at the end of the 2010–11 school year. Most of them (over 86%) were born in either 1992 or 1993. This cohort participated in the ALI incentive program for three years. We refer to this cohort as the class of 2011.

10th Grade in 2009-10 (Cohort 4 or class of 2012): These students began the first year of high school in the second year of the program. They would be expected to graduate in calendar year 2012, at the end of the 2011–12 school year. Most of them (over 87%) were born in either 1993 or 1994. This cohort experienced two years of ALI incentives (while in grades 10 and 11). We refer to this cohort as the class of 2012.

10th Grade in 2010-11 (Cohort 5 or class of 2013): These students began the first year of high school in the third year of the program. They would be expected to graduate in calendar year 2013, at the end of the 2012–13 school year. Most of them (over 88%) were born in either 1994 or 1995. This cohort received ALI incentives for one year (only while in 10th grade). We refer to this cohort as the class of 2013.

Cohorts 1 and 2 were in 11th and 12th grade at the time the ALI program was introduced, whereas cohorts 3, 4, and 5 participated in the ALI program starting in 10th grade. Given that some students drop out after 10th grade, cohorts 1 and 2 are generally more positively selected than cohorts 3, 4 and 5. This is not a source of bias, however, because the treatment and control groups are comparable across cohorts. Also, Behrman et al. (2015) found no evidence that the treatments affected grade progression, so the test scores of the treatment and control groups can be compared in subsequent years of the experiment. Of the five cohorts, we might expect the largest effects in cohort 3, which was exposed to the ALI program for three years.

Most of the participants were born between 1990 and 1995 (over 97% of the sample). The vast majority (99.9%) of students in the experimental sample turned 18 in 2013 or before. This means by

2019, the youngest of these students would be at least 24, by 2021, they would be at least 26, and by 2025, they would be at least 30. Figure A.1 shows the full distribution of birth years for each cohort.

4 Data

We rely on several datasets that allow us to identify individuals who participated in the ALI experiment (as treatments or controls) at different points in their lives. First, we use the ALI administrative records to recover the names and identifiers of all students enrolled in experimental schools.[6] The data collected by the Education Ministry (SEP) included the student's national identification number, the Unique Population Registry Code (*Clave Única de Registro de Población* or CURP), which can be used to track individual students over time (after they leave high school) in different datasets. The ALI data also include basic student demographics (state of birth, sex, and year of birth), the original randomized group assignment, and survey information used in the original studies that analyzed data from the experiment (Behrman et al., 2015; Todd & Wolpin, 2018).

We merge this information with the following databases:

- *Administrative test score data*: The National Assessment of Academic Achievement in Schools (*Evaluación Nacional del Logro Académico en Centros Escolares* or ENLACE) data are administrative student test score data on a nationwide standardized exam applied annually between 2006 and 2013 to nearly all students in Mexican public and private schools. ENLACE was a low-stakes test for students (including relative to the ALI program, as performance in ENLACE was not linked to any payments). The test was intended to measure educational quality, monitor policy effectiveness, and provide information to parents, teachers, and schools. The exam covered grades 3–6 and 9 every year. From 2008, it also included grade 12, and from 2009, it included grades 7–8. Students were tested in Spanish, mathematics, and one rotating subject. The test, administered over two days, consisted of 50–70 multiple-choice items per subject. The coverage is effectively universal, although Oaxaca and MichoacÃ¡n experienced years of missing or partial implementation due to opposition from the teachers' union. We use these

[6]There were 99,041 students enrolled in Grades 10–12 in experimental schools from 2008–09 to 2010–11. Of these, 54 changed schools during the experiment to another ALI school and were exposed to different treatments. We drop these cases from the data. Another 8 students changed schools within the experimental sample without changing treatment status, but changing randomization strata; we drop them as well. One student changed schools within the sample, kept the same treatment status, and the same block. We assign this student to their first school for clustering. Finally, 1,847 students had missing or incorrect identifiers (CURPS), and we excluded them because we were unable to match them to any other dataset (the proportion of missing/incorrect CURPS is balanced across experimental groups) — see Table A.1. As a result, our core sample comprises 97,132 students with valid CURPS.

data to retrieve students' scores in Grades 6-9, which are test scores prior to their participation in the ALI experiment. We use these variables primarily as control covariates to assess balance. We also compute the school's average ENLACE score in 2008, the year before the ALI program began. Finally, an indicator that equals 1 if the student took the grade 12 ENLACE is used as a proxy for high school graduation, as the test is administered near the end of 12th grade. We also use the ENLACE grade 12 scores as an academic achievement measure at the end of high school.

- *Tertiary degree data*: We compiled data from the registry of tertiary degrees (*cédulas profesionales*). A *cédula profesional* is the official license number issued by the Ministry of Education (Secretaría de Educación Pública, SEP). It certifies that a person has completed a recognized higher-education program (undergraduate or graduate) and is legally authorized to practice their profession. The registry is public and includes all degrees granted by accredited Mexican institutions.[7] We searched for all of the ALI participants in this public registry in October of 2025 and recorded whether they completed a tertiary degree, in what year, what type of degree they obtained, and in what field.[8]

- *Information on participation in public programs*: Through a research collaboration with the Federal Government of Mexico, we obtained data on whether students in the ENLACE dataset (which covers the vast majority of students enrolled in experimental schools from 2008 to 2011)benefited from any federal social programs. The programs include the conditional cash transfer program *Prospera*, the public health program *Seguro Popular*, and the public health program *IMSS-Bienestar*.

- *Formal-sector employment data*: Through a research collaboration with the Central Bank of Mexico (*Banco de México*), we obtained data from the Mexican Social Security Institute (*Instituto Mexicano del Seguro Social*, IMSS). These records contain detailed monthly data on social security contributions, wages, employment spells, and employer characteristics for the universe of private-sector formal workers in Mexico.

[7] The closest equivalent in the United States would be state-level professional licenses issued for regulated fields (e.g., law, medicine, engineering, teaching).

[8] The public registry is available at https://siurp.sep.gob.mx/mvc/cedulaElectronica and https://cedulaprofesional.sep.gob.mx/.

- *Voting registry data*: Through a research collaboration with the National Electoral Institute (*Instituto Nacional Electoral*, INE), we obtained voter registry data. These records indicate whether an individual voted in the 2015, 2018, and 2021 federal elections, along with self-reported data on educational attainment, labor force status, and residence characteristics recorded at the time of voter registration. In principle, individuals are expected to update their address when they move, which in turn updates their reported education and employment. In practice, however, few individuals update their voter registration, so these variables provide imperfect proxies for current characteristics.

5 Empirical specification

To estimate the long-term effects of the ALI incentives experiment, we exploit the randomized assignment of schools to the three treatment groups and to a control group. Our empirical strategy estimates intent-to-treat (ITT) effects by comparing student outcomes in three randomized treatment groups to those of students in the control group of schools.[9]

We estimate regression specifications of the form:

$$Y_{ics} = \sum_{c=1}^{5} \{\beta_{1c} T1_s + \beta_{2c} T2_s + \beta_{3c} T3_s\} + \gamma' X_{ic0} + \delta_{b(s)} + \delta_c + \varepsilon_{ics}, \quad (1)$$

where Y_{ics} is the outcome of interest for individual i in cohort c who attended school s, $T1_s$, $T2_s$, and $T3_s$ are indicator variables for whether school s was assigned to the treatments with student incentives, teacher incentives, or combined incentives. X_{ic0} includes baseline individual and school-level covariates (student demographics, ninth-grade ENLACE scores, and pre-ALI-program school characteristics). We also estimate the same equation without baseline controls, to confirm that the results are similar with and without covariate adjustments. The primary rationale for controlling for baseline covariates is to increase precision. The specification also includes fixed effects for the randomization strata ($\delta_{b(s)}$) and fixed effects for the student cohort (δ_c). Standard errors are clustered at the school level, which is the unit of randomization.

The coefficients of interest are β_{1c}, β_{2c}, and β_{3c}, which capture the cohort-specific ITT effects. Cohort 3 had the greatest exposure to the ALI program — three years (grades 10–12).

[9] As previously described, the sample of schools was first stratified into blocks and a block randomization was performed. Our regression specification, therefore, also includes block indicator variables.

The outcome measures we consider are an indicator of high school graduation, 12th-grade ENLACE test scores, an indicator of tertiary degree attainment, an indicator of whether the individual has a formal-sector job, formal-sector earnings, an indicator of social program participation, and an indicator of civic engagement.

Exposure to the treatment varied across students, both because different cohorts were potentially exposed to the treatment for different years (see Section 3.3) and because some students left treatment schools for different reasons (e.g., transfer to another school outside the experiment, drop out before graduating, or pause their studies temporarily). Therefore, we also estimate a two-stage least squares (2SLS) specification that uses treatment assignment interacted with student cohort as instruments for the number of semesters each individual was exposed to the different incentive regimes. Specifically, in the first stage, we regress the number of semesters of exposure to each treatment type on the interaction between the student's initial treatment assignment and the cohort, controlling for randomization strata, fixed effects, and cohort fixed effects. In the second stage, we estimate

$$Y_{is} = \alpha_1 \hat{D}_{is}^{(1)} + \alpha_2 \hat{D}_{is}^{(2)} + \alpha_3 \hat{D}_{is}^{(3)} + \gamma' X_{i0} + \delta_{b(s)} + \delta_{c(i)} + \varepsilon_{is}, \tag{2}$$

where $\hat{D}_{is}^{(k)}$ denotes the predicted number of semesters that individual i in school s was exposed to treatment $k \in \{1, 2, 3\}$. The excluded instruments are the interactions $T_s^{(k)} \times \mathbb{1}\{c(i) = c\}$.

Finally, we estimate all the treatment effects by cohort. This specification allows the treatment coefficients to vary flexibly across cohorts, capturing potential heterogeneity in program impacts arising from differences in exposure duration, baseline characteristics, and macroeconomic or policy environments at graduation. Specifically, we estimate versions of Equations (1) and (2) interacted with cohort dummies, thus identifying β_{1c}, β_{2c}, and β_{3c} for each cohort c.

5.1 Balance and attrition

In their analysis of the ALI experimental data, Behrman et al. (2015) showed that the observable characteristics of schools and students were balanced across experimental groups. In Tables A.2-A.3, we verify that the covariates are also balanced for students in our core sample; specifically, the percentage female, date of birth, Grade 9 ENLACE scores, and other time-invariant or pre-treatment characteristics are balanced across the experimental groups, both overall and within cohorts. We also

examined whether students across the experimental groups were equally likely to be present in the INE and Social Programs databases. Overall, the matching rate between the set of students enrolled in ALI experimental schools from 2008 to 2011 and the INE and Social Programs databases is high (over 95%), and similar across the experimental groups.Our tertiary-degree dataset contains only individuals who obtain tertiary degrees, whereas the IMSS data contain information on individuals with a formal-sector job. Therefore, our outcome variables for these domains are indicators of whether individuals are present in those datasets.

Our analysis assumes that taking the ENLACE 12 exam is a proxy for high school graduation. The set of students for whom we observe a 12th-grade ENLACE score is a subsample of all ALI participants. However, all students in the ALI program completed 9th grade. We examined whether the proportion of students with 9th-grade test scores in the ENLACE data is balanced across the experimental groups and find that it is (Tables A.2–A.3). Because T3 statistically significantly reduces the probability of taking the Grade 12 ENLACE exam (Section 6.1), conditional score estimates for this arm may be affected by selection. We note this limitation when interpreting those results.

All binary outcome variables are scaled from 0 to 100, so that regression coefficients can be read directly as percentage-point effects.

6 Results

6.1 Short-run academic outcomes at the end of high school

We first examine outcomes measured at the end of high school.[10] The likelihood of ever taking the Grade 12 ENLACE decreases by 3.6–4.2 percentage points (from a 70% base) for students in the combined-incentive group (T3; Table 3, Columns 1–2). The effect is negative for the other treatment arms but statistically insignificant. This pattern is driven by all cohorts except the first, which was already in Grade 12 when the program began (Figure 1). However, the effect does not appear to be driven by lower-performing or high-performing students (Table A.4).

[10]The following facts, outlined by Dustan, de Janvry, and Sadoulet (2017), suggest that most students who take the ENLACE 12 exam graduate from high school: (1) most dropout occurs before Grade 12 —of students who drop out, only 13–18% do so after advancing to 12th grade; (2) the average ENLACE registration rate is 99% of all 12th-grade students; (3) the average ENLACE taking rate is 98% of graduating 12th graders; and (4) only 0.25% of ENLACE-taking students retake the exam. See Dustan et al. (2017) for a detailed discussion.

Conditional on test-taking, the effect on mathematics scores is negative and statistically significant for the student-incentive arm (T1), consistent with the original findings from Behrman et al. (2015).[11]

Both findings — the decline in ENLACE 12 test-taking for the combined-incentive arm and the negative math scores for the student-incentive arm — are puzzling in light of the original short-run gains on the incentivized ALI exams (Behrman et al., 2015). One possible explanation is that ENLACE was a low-stakes exam relative to the ALI assessments, which directly determined bonus payments. The salience of incentives and the testing environment can affect effort and measured performance (Levitt et al., 2016; Gneezy et al., 2017). A series of teacher-incentive studies has consistently shown that treatment effects on incentivized tests exceed those on non-incentivized tests (Mbiti et al., 2019, 2023). More broadly, the theoretical literature on intrinsic and extrinsic motivation suggests that external rewards can crowd out internal motivation under certain conditions (Gneezy & Rustichini, 2000; Bénabou & Tirole, 2003).

[11] The positive point estimate on math scores for the combined-incentive arm may reflect changes in the composition of test-takers, given the negative effect on test-taking rates.

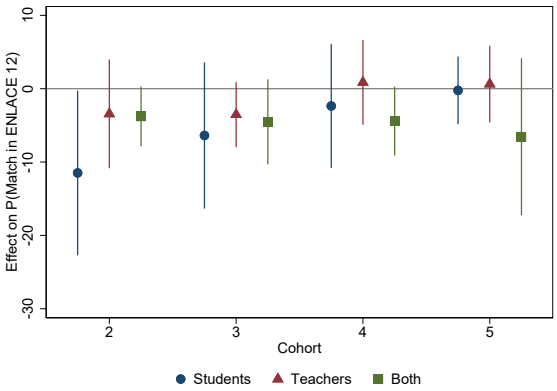
Table 3: Effect on ENLACE 12 outcomes

	Has Grade 12 ENLACE		Spanish score (Grade 12)		Math score (Grade 12)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Intent-to-treat						
Students	-3.94	-1.84	-0.12	-0.08	-0.50***	-0.53***
	(3.06)	(1.93)	(0.11)	(0.11)	(0.16)	(0.13)
Teachers	-2.03	-1.42	-0.13	-0.12	-0.47***	-0.46***
	(2.51)	(1.99)	(0.08)	(0.08)	(0.16)	(0.13)
Both	-4.15**	-3.60**	0.01	-0.01	-0.21	-0.23
	(1.90)	(1.75)	(0.09)	(0.09)	(0.15)	(0.14)
Control mean	70	70	.1	.1	.92	.92
p-value equality	.71	.5	.28	.34	.088	.057
R^2	.072	.11	.0016	.0062	.0014	.0031
N. of obs.	97,176	97,176	66,076	66,076	66,076	66,076
Panel B: LATE						
Semesters exposed to student incentives	-1.39	-0.85	-0.03	-0.02	-0.10***	-0.11***
	(1.04)	(0.69)	(0.02)	(0.02)	(0.03)	(0.03)
Semesters exposed to teacher incentives	-0.60	-0.45	-0.03	-0.02	-0.10***	-0.10***
	(0.67)	(0.52)	(0.02)	(0.01)	(0.03)	(0.03)
Semesters exposed to both incentives	-1.20**	-1.00**	0.01	0.01	-0.02	-0.03
	(0.52)	(0.49)	(0.02)	(0.02)	(0.03)	(0.03)
Control mean	70	70	.1	.1	.92	.92
F-stat (first stage)	2,879	2,706	37,204	33,362	37,204	33,362
p-value equality	.66	.56	.18	.29	.036	.019
R^2	-.016	.036	.000089	.0047	.00017	.0019
N. of obs.	97,176	97,176	66,076	66,076	66,076	66,076
Controls	No	Yes	No	Yes	No	Yes

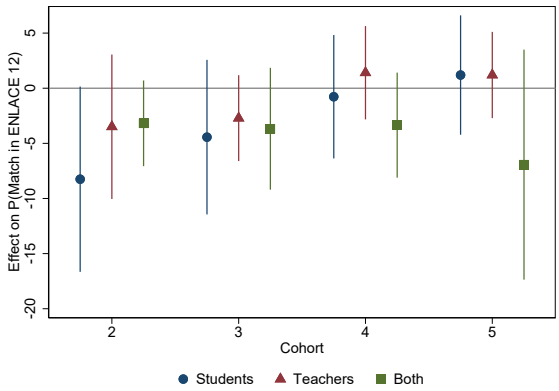
Notes: All specifications include strata and cohort fixed effects. Panel A reports intent-to-treat (ITT) estimates following Equation 1; Panel B reports local average treatment effects (LATE) following Equation 2 estimated via two-stage least squares. The outcome in Columns 1–2 is whether a student ever takes the ENLACE Grade 12 exam (0/100); Columns 3–4 report Spanish scores (conditional on test-taking); and Columns 5–6 report mathematics scores (conditional on test-taking). When a student takes the exam multiple times, we retain the highest observed score. Odd-numbered columns include no additional controls. Even-numbered columns add: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Treatment effects by cohort on ENLACE 12 outcomes

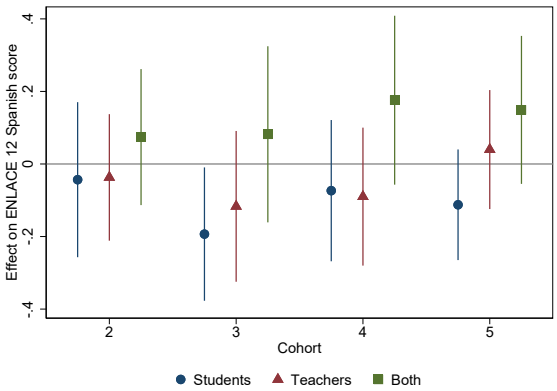
(a) P(taking the ENLACE 12)



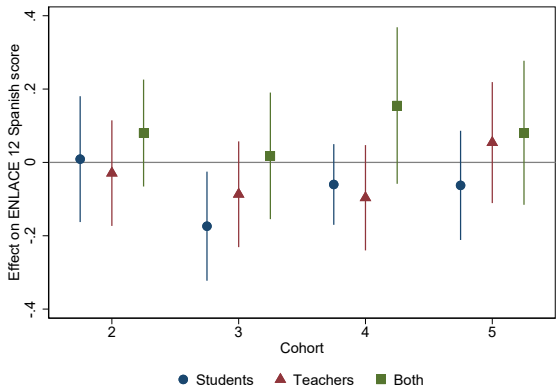
(b) P(taking the ENLACE 12) with controls



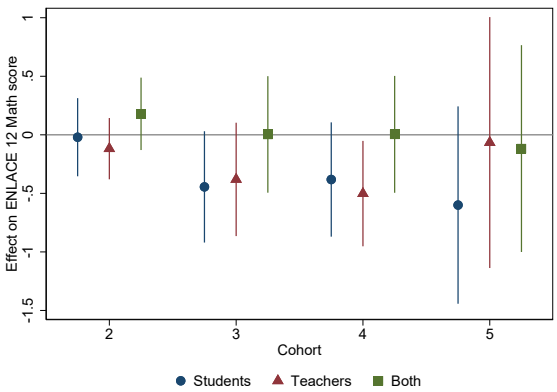
(c) ENLACE 12 Spanish score



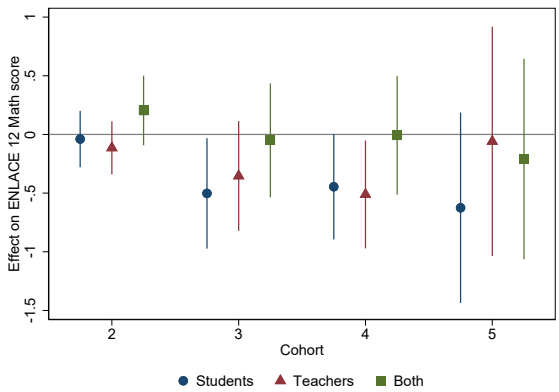
(d) ENLACE 12 Spanish score with controls



(e) ENLACE 12 math score



(f) ENLACE 12 math score with controls



Note: Each panel plots estimated treatment effects by cohort for the probability of taking the ENLACE Grade 12 exam and for Spanish and mathematics scores (conditional on test-taking). Point estimates and 95% confidence intervals are shown for each treatment arm (student incentives, teacher incentives, and combined incentives). Figures on the left present specifications without additional controls; figures on the right include the following controls: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. All regressions include randomization–strata and cohort fixed effects, and standard errors are clustered at the school level.

6.2 Professional degrees

The schools in our experiment belong to one of two subsystems of federal upper-secondary education in Mexico: the *Dirección General de Educación Tecnológica Industrial* (DGETI) and the *Dirección General de Educación Tecnológica Agropecuaria* (DGETA). Both types of institutions award technical degrees in addition to the standard high school diploma, so some students may already have earned a postsecondary credential upon graduation.

Table 4 reports effects on the probability of holding a tertiary degree by 2025, distinguishing between technical, bachelor's, and field-specific degrees. Students assigned to the student-incentive arm (T1) were 4–5 percentage points more likely to hold any post-secondary degree, relative to a 22% baseline in control schools. Most of the effect is attributable to bachelor's degree attainment.However, the teacher-incentive (T2) and combined-incentive (T3) arms show no significant effects; we can rule out effects larger than approximately 5 percentage points for T2 and T3 at the 95% level. Although the ALI incentives targeted mathematics, we find no effect on the likelihood of obtaining a STEM-related degree or a degree in economics or business.

The effect is driven primarily by the first four cohorts and approaches zero for the fifth cohort, which received only one year of exposure starting in 10th grade (Figure 2).

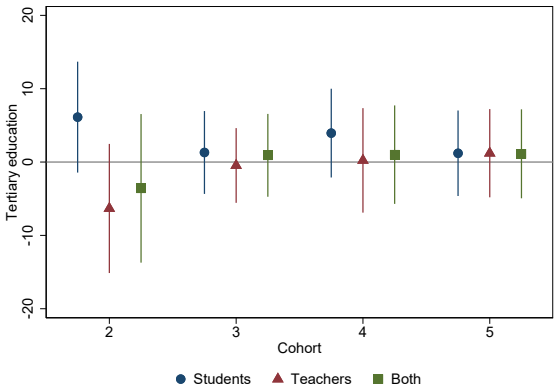
Table 4: Effect on the likelihood of having a post-secondary education degree by 2025

	Any post-secondary degree		Technical degree		Bachelor's degree		STEM degree		Econ/business degree	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Intent-to-treat										
Students	3.87*	5.01**	0.59	0.51	-0.01	-0.00	1.02	1.79	0.15	0.59
	(2.28)	(1.98)	(0.45)	(0.43)	(0.00)	(0.00)	(1.54)	(1.57)	(0.99)	(0.86)
Teachers	-0.76	-0.72	0.10	0.08	-0.01	-0.01	0.99	0.96	-0.82	-0.75
	(2.20)	(1.87)	(0.35)	(0.36)	(0.00)	(0.00)	(1.33)	(1.24)	(0.91)	(0.82)
Both	-0.03	0.12	-0.30	-0.40	-0.01	-0.01	2.28	2.03	0.36	0.15
	(1.96)	(1.80)	(0.22)	(0.26)	(0.00)	(0.00)	(1.48)	(1.36)	(0.95)	(0.85)
Control mean	22	22	.67	.67	.0065	.0065	10	10	4.9	4.9
p-value equality	.16	.0059	.06	.07	.76	.76	.67	.72	.36	.18
R^2	.058	.098	.0073	.0089	.00027	.00038	.031	.055	.014	.026
N. of obs.	97,176	97,176	97,176	97,176	97,176	97,176	97,176	97,176	97,176	97,176
Panel B: LATE										
Semesters exposed to student incentives	1.00	1.28**	0.14	0.11	-0.00	-0.00	0.23	0.43	0.01	0.12
	(0.64)	(0.56)	(0.13)	(0.12)	(0.00)	(0.00)	(0.45)	(0.44)	(0.29)	(0.25)
Semesters exposed to teacher incentives	-0.31	-0.30	0.02	0.02	-0.00	-0.00	0.25	0.24	-0.29	-0.27
	(0.61)	(0.53)	(0.11)	(0.12)	(0.00)	(0.00)	(0.38)	(0.36)	(0.26)	(0.24)
Semesters exposed to both incentives	-0.03	0.04	-0.11	-0.13*	-0.00	-0.00	0.64	0.58	0.05	-0.01
	(0.60)	(0.56)	(0.07)	(0.08)	(0.00)	(0.00)	(0.46)	(0.41)	(0.28)	(0.25)
Control mean	22	22	.67	.67	.0065	.0065	10	10	4.9	4.9
F-stat (first stage)	2,954	2,681	2,954	2,681	2,954	2,681	2,954	2,681	2,954	2,681
p-value equality	.15	.007	.059	.066	.75	.78	.64	.69	.34	.16
R^2	.0035	.046	.0012	.0028	.000033	.00014	.0032	.028	.00011	.012
N. of obs.	97,176	97,176	97,176	97,176	97,176	97,176	97,176	97,176	97,176	97,176
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

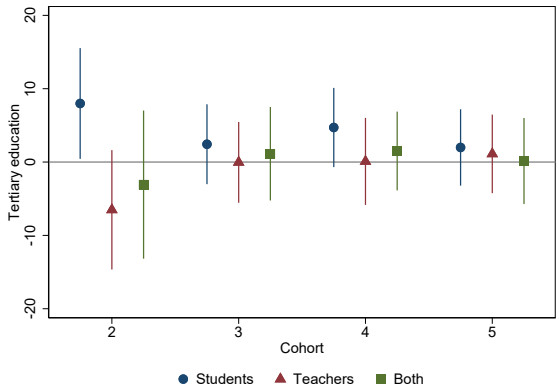
Notes: All specifications include strata and cohort fixed effects. Panel A reports intent-to-treat (ITT) estimates following Equation 1; Panel B reports local average treatment effects (LATE) following Equation 2 estimated via two-stage least squares. The outcome in Columns 1–2 is whether a student has any post-secondary degree (0/100); Columns 3–4 whether the student holds a technical degree; Columns 5–6 whether the student holds a Bachelor's degree; Columns 7–8 whether the student holds a degree in a STEM field, and Columns 9–10 whether the student holds a degree in an economics or business related field. Odd-numbered columns include no additional controls. Even-numbered columns add: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2: Treatment effects by cohort on the likelihood of holding a post-secondary degree

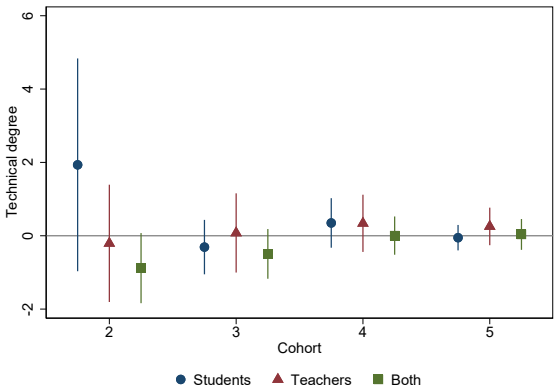
(a) Any post-secondary degree



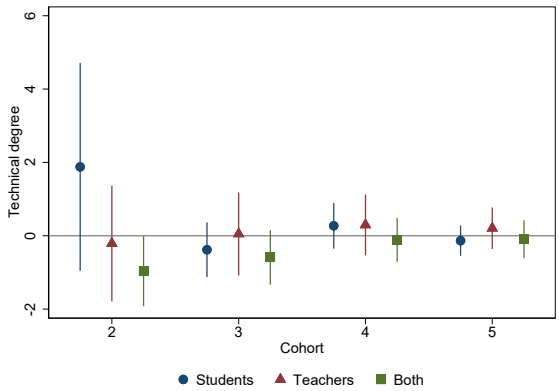
(b) Any post-secondary degree with controls



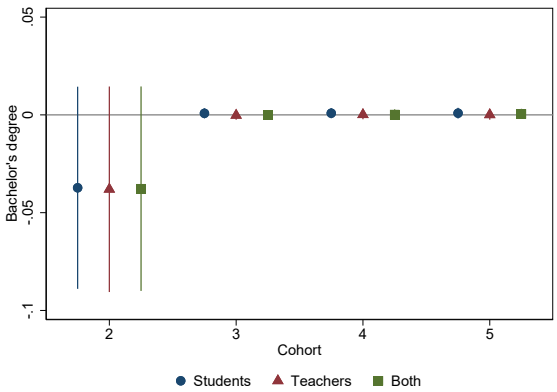
(c) Technical degree



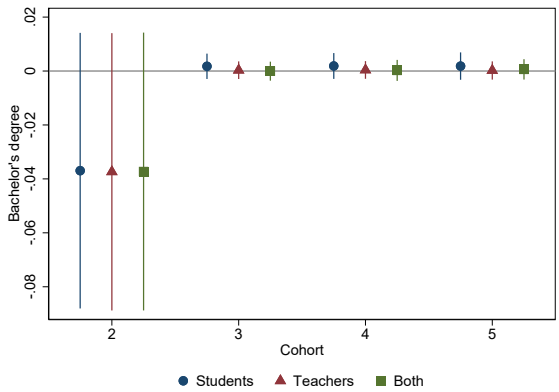
(d) Technical degree with controls



(e) Bachelor's degree



(f) Bachelor's degree with controls



Note: Each panel plots estimated treatment effects by cohort for the probability of holding a post-secondary degree (any degree, technical degree, or bachelor's degree). Point estimates and 95% confidence intervals are shown for each treatment arm (student incentives, teacher incentives, and combined incentives). Figures on the left present specifications without additional controls; figures on the right include the following controls: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. All regressions include randomization-strata and cohort fixed effects, and standard errors are clustered at the school level.

6.3 Formal employment

We use monthly administrative records from the Mexican Social Security Institute (*Instituto Mexicano del Seguro Social*, IMSS) to examine whether exposure to ALI incentives affected formal-sector labor market outcomes. These records cover the universe of private-sector formal employment in Mexico and include monthly data on social security contributions, wages, employment spells, and employer characteristics (industry, firm size, and location). We access these data through a research collaboration with the Central Bank of Mexico (*Banco de México*), conducted inside the Bank's Economic Microdata Laboratory (*EconLab*).[12]

We examine four outcomes: (i) an indicator for ever appearing in formal employment records, (ii) an indicator for holding a formal-sector job in the last year, (iii) log monthly wages conditional on formal employment, and (iv) the cumulative number of months in formal employment. The identification strategy mirrors the ITT and 2SLS specifications described in Equations 1 and 2, with standard errors clustered at the school level.

Preliminary point estimates suggest that the student-incentive arm (T1) increases the probability of holding a formal-sector job. This finding is consistent with the reduction in enrollment in *IMSS-Bienestar* and *Seguro Popular* documented in Section 6.4, as these programs serve individuals without formal-sector social security coverage. However, these results should be interpreted with caution pending final extraction and verification from the EconLab. We expect to incorporate the complete formal employment analysis before the NBER Summer Institute presentation.

6.4 Social programs

Next, we examine whether individuals who attended treatment schools were more or less likely to be beneficiaries of major social programs in August of 2019. In Mexico, the largest social assistance programs during the study period were *Prospera* (formerly *Oportunidades*), *Seguro Popular*, and *IMSS-Bienestar*. *Prospera* was a conditional cash transfer program that provided income support to poor households conditional on school attendance, health checkups, and nutrition requirements. *Seguro Popular* was a public health insurance scheme designed to extend access to medical services to individuals without formal employment or social security

[12]The EconLab is a secure computing environment at the Central Bank of Mexico that provides researchers access to IMSS administrative records. All analyses are conducted in the laboratory, and output extraction requires approval by laboratory staff, with a minimum of 5 observations per cell. See Banco de México (2023) for details on access protocols.

coverage. *Seguro Popular* was replaced by *IMSS-Bienestar*, managed by the Mexican Social Security Institute (IMSS). We use administrative records that link ALI participants to federal program beneficiary registries to estimate whether exposure to the student, teacher, or combined incentive schemes affected later participation in these programs.

Assignment to ALI treatment schools reduced students' later participation in social assistance programs (Table 5; Figure 3). Students in the student-incentive arm (T1) were approximately 7 percentage points less likely to appear in any major federal welfare registry relative to the 57% control mean, with the reduction driven by a decline in enrollment in *IMSS-Bienestar* or *Seguro Popular*. Teacher incentives (T2) and combined incentives (T3) show smaller, statistically insignificant declines. The 2SLS results (Panel B) show a similar pattern.

Because *IMSS-Bienestar* and *Seguro Popular* cover individuals without formal-sector employment, these declines are consistent with treated students being more likely to hold formal-sector jobs with social security coverage in adulthood. However, we cannot rule out that changes in eligibility criteria, take-up behavior, or geographic mobility to areas with different program coverage also contribute to the observed differences. Program enrollment is observed at a single point (August 2019) during a period of major administrative restructuring of social programs in Mexico, which may affect both coverage and measurement.

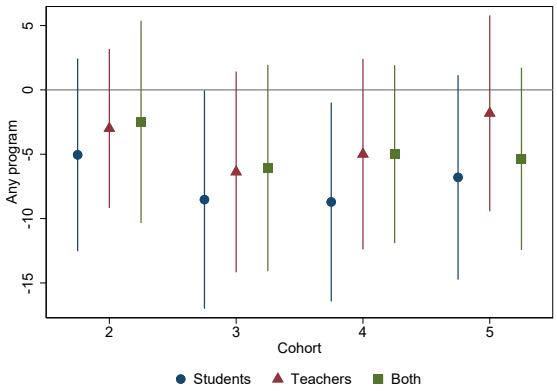
Table 5: Effect on benefits from a social program

	Any program		Prospera		Seguro Popular or IMSS-B	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Intent-to-treat						
Students	-7.24*	-4.05	-2.04	-0.47	-7.05**	-3.69
	(3.79)	(3.44)	(2.83)	(2.45)	(3.20)	(3.35)
Teachers	-3.95	-3.13	-0.32	0.07	-4.76	-3.91
	(3.35)	(3.00)	(2.75)	(2.25)	(3.01)	(2.64)
Both	-4.66	-3.82	-1.37	-1.36	-3.62	-2.82
	(3.61)	(3.03)	(2.36)	(2.00)	(3.15)	(2.82)
Control mean	57	57	16	16	44	44
p-value equality	.69	.96	.87	.79	.62	.93
R^2	.052	.069	.021	.039	.043	.1
N. of obs.	85,639	85,639	85,639	85,639	97,176	97,176
Panel B: LATE						
Semesters exposed to student incentives	-2.05*	-1.17	-0.55	-0.12	-2.03**	-1.08
	(1.07)	(0.95)	(0.77)	(0.66)	(0.89)	(0.90)
Semesters exposed to teacher incentives	-1.22	-1.00	-0.07	0.03	-1.40*	-1.21
	(0.94)	(0.83)	(0.75)	(0.61)	(0.83)	(0.73)
Semesters exposed to both incentives	-1.34	-1.11	-0.34	-0.32	-1.12	-0.88
	(1.04)	(0.85)	(0.67)	(0.55)	(0.89)	(0.79)
Control mean	57	57	16	16	44	44
F-stat (first stage)	3,277	3,084	3,277	3,084	2,879	2,706
p-value equality	.73	.99	.86	.83	.64	.91
R^2	.0032	.021	.00033	.018	.0021	.062
N. of obs.	85,639	85,639	85,639	85,639	97,176	97,176
Controls	No	Yes	No	Yes	No	Yes

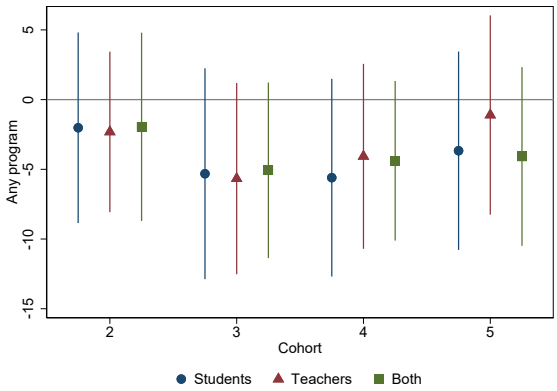
Notes: All specifications include strata and cohort fixed effects. Panel A reports intent-to-treat (ITT) estimates following Equation 1; Panel B reports local average treatment effects (LATE) following Equation 2, estimated via two-stage least squares. The dependent variable in Columns 1–2 indicates whether a student benefits from any federal social program (0/100); Columns 3–4 correspond to *Prospera* (formerly *Oportunidades*); and Columns 5–6 to *Seguro Popular* or *IMSS-Bienestar*. Odd-numbered columns include no additional controls. Even-numbered columns add: school-level Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3: Treatment effects by cohort on the likelihood of benefiting from social programs

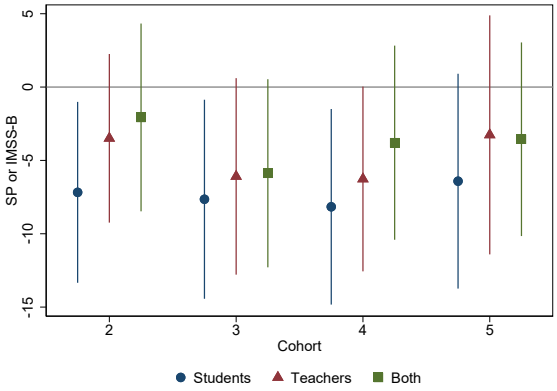
(a) Any social program



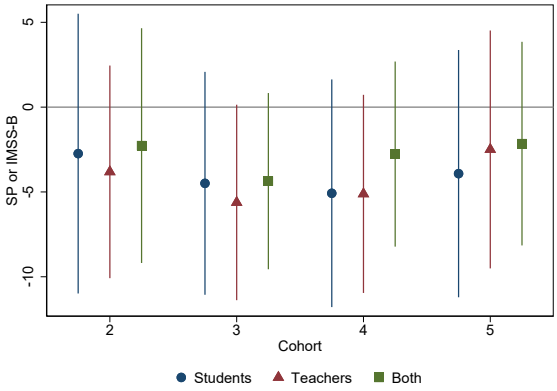
(b) Any social program with controls



(c) Seguro Popular or IMSS-B



(d) Seguro Popular or IMSS-B with controls



Note: Each panel plots estimated treatment effects by cohort for the probability of benefiting from a federal social program (any program or Seguro Popular/IMSS-Bienestar specifically). Point estimates and 95% confidence intervals are shown for each treatment arm (student incentives, teacher incentives, and combined incentives). Figures on the left present specifications without additional controls; figures on the right include the following controls: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. All regressions include randomization-strata and cohort fixed effects, and standard errors are clustered at the school level.

6.5 Electoral institute data

We use administrative data from the National Electoral Institute (*Instituto Nacional Electoral*, INE) to examine effects on civic participation and geographic mobility. In Mexico, the INE issues the national voting ID (*Credencial para Votar*), which is required to vote and serves as the country's primary form of official identification across public and private transactions. The INE also provides an online credential-verification service used by banks and other institutions. When a new ID is issued, individuals must report their place of residence to be assigned to a polling station in future

elections. In principle, individuals are expected to update their addresses when they move, allowing us to merge reported localities with census data to characterize the socioeconomic attributes of each participant's community.[13] Beyond information on ID issuance and residence, we observe whether participants voted in the 2018 and 2021 federal elections. The 2018 election included the presidency and registered a turnout of 63.4%, whereas the 2021 midterm (legislative) election recorded a turnout of 52.7% (Instituto Nacional Electoral, 2018, 2021).

We find no meaningful effect on the likelihood of holding a voting ID by September 2021, with possession rates near 86% in the control group (Table 6; Figure 4). The combined-incentive arm (T3) shows a modest 1.6 percentage-point increase once controls are included ($p < 0.10$). However, voter turnout falls under the same arm: participation in the 2018 presidential election declines by about 4 percentage points from a 66% base, and the 2SLS estimates imply a reduction of roughly 1 percentage point per semester of exposure. Student-only (T1) and teacher-only (T2) incentives have negligible effects on turnout. In the 2021 midterm, point estimates for T3 remain negative (-4.4 to -2.8 pp), but are less precise.

[13] In practice, many individuals do not update their registration, so this variable provides only an imperfect proxy for current residence.

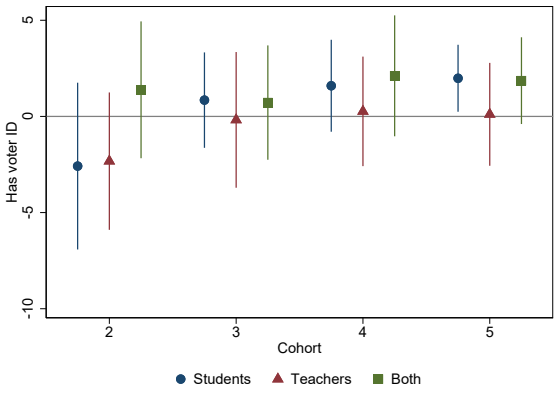
Table 6: Effect on civic duties

	Voting ID (as of Sep 2021)		Voted in 2018		Voted in 2021	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Intent-to-treat						
Students	0.94	0.78	0.86	1.79	-1.15	0.02
	(0.92)	(0.91)	(1.65)	(1.48)	(1.83)	(1.98)
Teachers	-0.61	-0.74	0.60	0.91	-2.07	-2.15
	(1.30)	(0.92)	(1.77)	(1.51)	(2.12)	(1.75)
Both	1.69	1.56*	-4.34**	-3.13*	-4.36**	-2.80
	(1.14)	(0.93)	(1.68)	(1.61)	(1.94)	(1.73)
Control mean	86	86	66	66	59	59
p-value equality	.26	.026	.0064	.016	.22	.32
R^2	.03	.27	.013	.033	.016	.045
N. of obs.	97,176	97,176	97,176	97,176	97,176	97,176
Panel B: LATE						
Semesters exposed to student incentives	0.18	0.14	0.25	0.48	-0.29	0.02
	(0.27)	(0.26)	(0.45)	(0.40)	(0.52)	(0.55)
Semesters exposed to teacher incentives	-0.16	-0.27	0.12	0.22	-0.55	-0.54
	(0.36)	(0.25)	(0.49)	(0.41)	(0.57)	(0.47)
Semesters exposed to both incentives	0.44	0.44	-1.22**	-0.88*	-1.19**	-0.76
	(0.34)	(0.28)	(0.49)	(0.46)	(0.56)	(0.49)
Control mean	86	86	66	66	59	59
F-stat (first stage)	2,879	2,706	2,879	2,706	2,879	2,706
p-value equality	.32	.0075	.008	.017	.24	.34
R^2	.0014	.25	.0016	.023	.0012	.03
N. of obs.	97,176	97,176	97,176	97,176	97,176	97,176
Controls	No	Yes	No	Yes	No	Yes

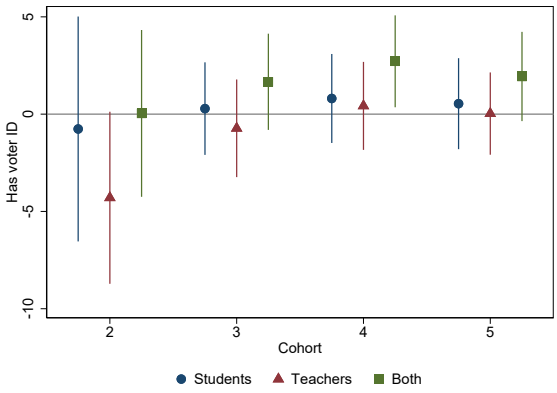
Notes: All specifications include strata and cohort fixed effects. Panel A reports intent-to-treat (ITT) estimates following Equation 1; Panel B reports local average treatment effects (LATE) following Equation 2, estimated via two-stage least squares. The dependent variable in Columns 1-2 indicates whether the individual holds a valid voting ID (*Credencial para Votar*) as of September 2021, according to administrative records from the *Instituto Nacional Electoral* (INE). Columns 3–4 report whether the individual voted in the 2018 presidential election, and Columns 5–6 whether they voted in the 2021 midterm legislative election. Odd-numbered columns include no additional controls. Even-numbered columns add: school-level Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 4: Treatment effects by cohort on the likelihood of having a voting ID and voting

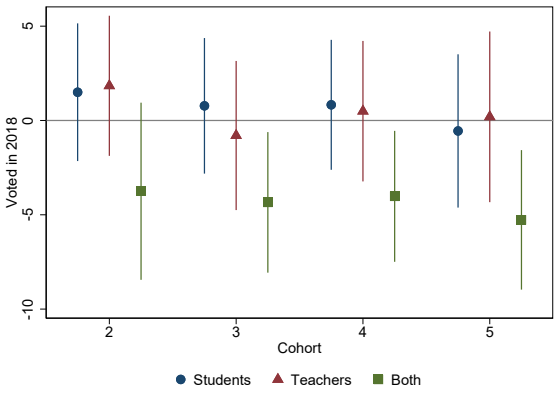
(a) Voting ID



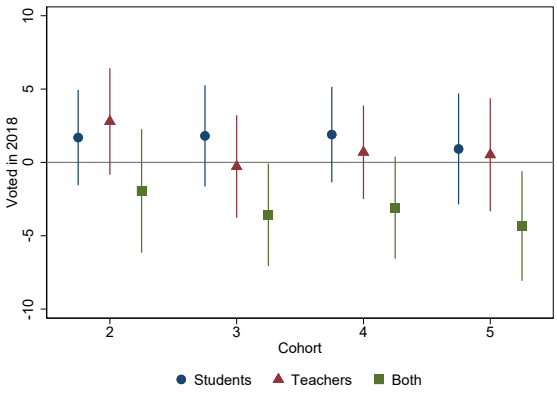
(b) Voting ID with controls



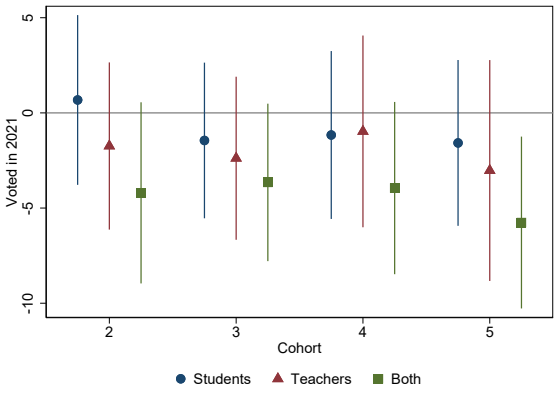
(c) Voted 2018



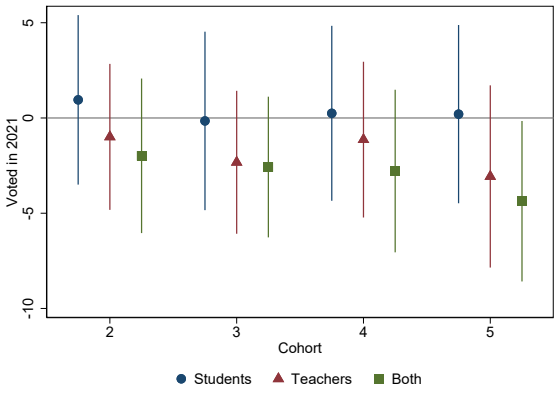
(d) Voted 2018 with controls



(e) Voted 2021



(f) Voted 2021 with controls



Note: Each panel plots estimated treatment effects by cohort for the probability of holding a voting ID (Credencial para Votar) and for voting in the 2018 and 2021 federal elections. Point estimates and 95% confidence intervals are shown for each treatment arm (student incentives, teacher incentives, and combined incentives). Figures on the left present specifications without additional controls; figures on the right include the following controls: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. All regressions include randomization-strata and cohort fixed effects, and standard errors are clustered at the school level.

We next examine the characteristics of the localities where ALI participants reside, based on the address reported on their latest voting ID as of September 2021 (Table 7; Figure 5). Participants exposed to student (T1) or combined (T3) incentives live in localities with larger populations than control peers. However, neither unemployment rates, education levels, nor the marginalization index show systematic differences. These patterns are consistent with either greater geographic mobility toward more urban areas or higher rates of address updating among treated individuals (e.g., because formal employment requires a current ID). The imprecision of the address-updating mechanism limits the interpretation of this finding.

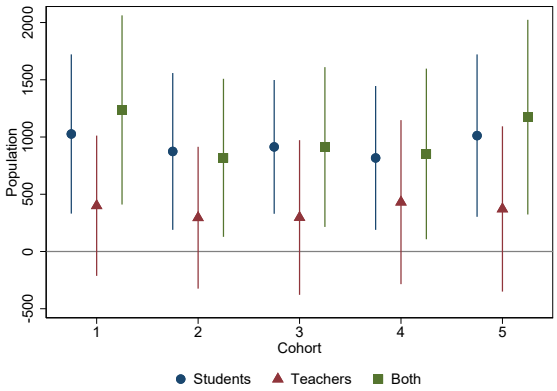
Table 7: Effect on the characteristics of where people live

	Population		Years of schooling		Unemployment rate		Marginalization index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Intent-to-treat								
Students (T1)	923.09***	861.20***	0.28	0.21	-0.03	-0.20	0.00	0.00
	(319.07)	(293.80)	(0.18)	(0.15)	(0.15)	(0.16)	(0.01)	(0.01)
Teachers (T2)	358.97	244.45	-0.01	0.00	0.23	0.19	0.00	0.01
	(332.69)	(215.37)	(0.16)	(0.16)	(0.22)	(0.18)	(0.01)	(0.01)
Both (T3)	988.64***	907.92**	0.14	0.19	0.37	0.37*	0.01	0.00
	(373.16)	(376.39)	(0.18)	(0.17)	(0.23)	(0.20)	(0.01)	(0.01)
Control mean	2,864	2,864	9	9	1.7	1.7	.88	.88
p-value equality	.25	.04	.38	.43	.18	.025	.9	.96
R^2	.023	.035	.017	.04	.014	.032	.026	.052
N. of obs.	83,056	83,056	83,056	83,056	83,038	83,038	66,592	66,592
Panel B: LATE								
Semesters exposed to T1	245.78***	226.27***	0.08	0.06	-0.01	-0.05	0.00	0.00
	(84.24)	(75.62)	(0.05)	(0.04)	(0.04)	(0.04)	(0.00)	(0.00)
Semesters exposed to T2	93.98	63.43	-0.00	-0.00	0.06	0.05	0.00	0.00
	(89.11)	(57.09)	(0.04)	(0.04)	(0.06)	(0.05)	(0.00)	(0.00)
Semesters exposed to T3	261.46**	236.36**	0.04	0.05	0.10	0.10*	0.00	0.00
	(100.82)	(101.48)	(0.05)	(0.05)	(0.06)	(0.05)	(0.00)	(0.00)
Control mean	2,864	2,864	9	9	1.7	1.7	.88	.88
F-stat (first stage)	3,462	3,032	3,462	3,032	3,463	3,039	3,495	3,030
p-value equality	.25	.041	.35	.38	.18	.026	.9	.97
R^2	.0051	.017	.0051	.029	.0069	.024	.00067	.027
N. of obs.	83,056	83,056	83,056	83,056	83,038	83,038	66,592	66,592
Controls	No	Yes	No	Yes	No	Yes	No	Yes

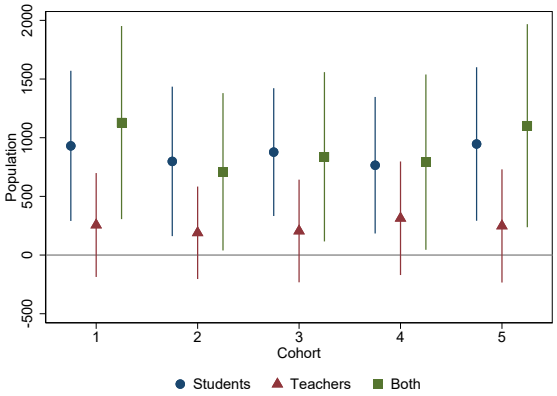
Notes: All specifications include strata and cohort fixed effects. Panel A reports intent-to-treat (ITT) estimates following Equation 1; Panel B reports local average treatment effects (LATE) following Equation 2, estimated via two-stage least squares. The dependent variables describe characteristics of the locality where each participant resides as of September 2021, based on their registered address in the *Instituto Nacional Electoral* (INE) database. Columns 1–2 report the total population of the municipality (in hundreds of residents); Columns 3–4 the average years of schooling among adults; Columns 5–6 the local unemployment rate; and Columns 7–8 the marginalization index (standardized at the national level). Odd-numbered columns include no additional controls. Even-numbered columns add: school-level Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5: Treatment effects by cohort on the characteristics of the location of residence

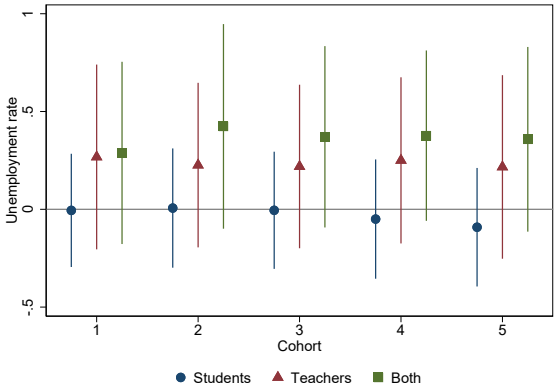
(a) Population



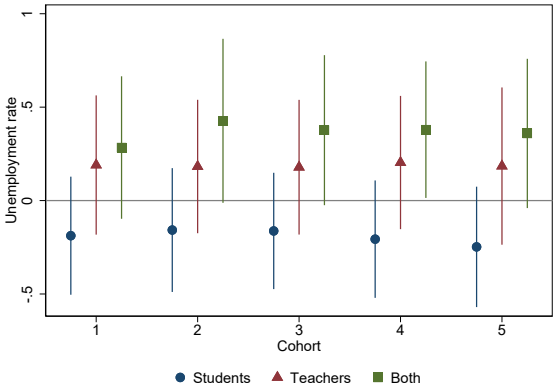
(b) Population with controls



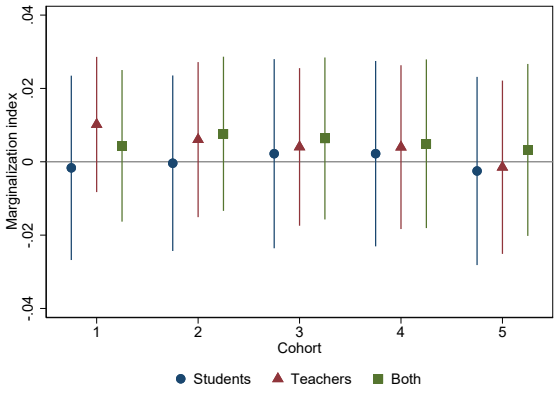
(c) Unemployment rate



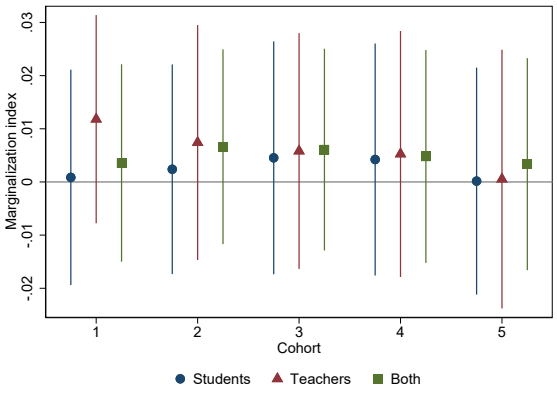
(d) Unemployment rate with controls



(e) Marginalization index



(f) Marginalization index with controls



Note: Each panel plots estimated treatment effects by cohort for characteristics of the locality where participants reside as of September 2021 (population, unemployment rate, and marginalization index). Point estimates and 95% confidence intervals are shown for each treatment arm (student incentives, teacher incentives, and combined incentives). Figures on the left present specifications without additional controls; figures on the right include the following controls: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), gender, year of birth, an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. All regressions include randomization-strata and cohort fixed effects, and standard errors are clustered at the school level.

7 Conclusion

This paper evaluates the long-term effects of a large-scale randomized incentive program — the Aligning Learning Incentives (ALI) experiment — that paid students, teachers, or both based on mathematics test performance in 88 Mexican public high schools. By linking the original experimental data with national administrative records spanning education, social programs, voter rolls, and formal employment, we trace the effects of performance incentives over a 10–15 year horizon.

Our analysis yields four main findings. First, the short-run test-score gains documented by Behrman et al. (2015) on the incentivized ALI exams did not carry over to the low-stakes ENLACE Grade 12 exam. Students in the combined-incentive arm were approximately 4 percentage points less likely to take the Grade 12 ENLACE (a proxy for high school completion), and conditional test scores did not improve. This pattern is consistent with evidence that incentive effects are strongest on directly incentivized assessments and may not transfer to low-stakes testing environments. Second, student incentives increased the probability of obtaining a post-secondary degree by roughly 5 percentage points relative to a 22% baseline — an effect driven primarily by bachelor's degree attainment and concentrated among the first four cohorts. Third, student incentives reduced enrollment in federal social assistance programs by approximately 4–7 percentage points (depending on the inclusion of controls), with the decline concentrated in *IMSS-Bienestar* and *Seguro Popular* — programs that serve individuals without formal-sector employment. This finding is consistent with higher rates of formal-sector labor-market participation among treated students, although the evidence on formal employment remains preliminary. Fourth, the combined student-and-teacher incentive arm reduced voter turnout by about 4 percentage points in the 2018 presidential election, and student and combined incentives are associated with residence in larger, less marginalized localities.

From a policy perspective, our findings suggest that the ALI student incentive program generated lasting returns on human capital in this setting, even when the direct test-score effects are concentrated on incentivized assessments. The reduction in social program dependence is economically meaningful, and preliminary evidence of a shift toward formal employment would further strengthen this interpretation if confirmed. The negative effects on voter turnout in the

combined-incentive arm, though modest in magnitude, remain unexplained — we do not have a compelling mechanism for why combined incentives would reduce turnout while student-only and teacher-only incentives do not — and should be treated as exploratory pending replication.

References

Banco de México. (2023). *Reglamento para el uso de los servicios del Laboratorio de Microdatos Económicos (EconLab).* https://www.banxico.org.mx/.

Barrera-Osorio, F., & Raju, D. (2017). Teacher performance pay: Experimental evidence from pakistan. *Journal of Public Economics*, *148*, 75-91. Retrieved from https://www.sciencedirect.com/science/article/pii/S0047272717300142 doi: https://doi.org/10.1016/j.jpubeco.2017.02.001

Behrman, J. R., Parker, S. W., Todd, P. E., & Wolpin, K. I. (2015). Aligning learning incentives of students and teachers: Results from a social experiment in mexican high schools. *Journal of Political Economy*, *123*(2), 325–364. Retrieved 2025-09-26, from http://www.jstor.org/stable/10.1086/675910

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, *70*(3), 489–520.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679.

Cohodes, S., Eren, O., & Ozturk, O. (2023, March). *The long run effects of a comprehensive teacher performance pay program on student outcomes* (Working Paper No. 31056). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w31056 doi: 10.3386/w31056

Cohodes, S. R., Eren, O., & Ozturk, O. (2026). The long run effects of a teacher-focused school reform on student outcomes. *Journal of Public Economics*, *253*, 105561.

Dustan, A., de Janvry, A., & Sadoulet, E. (2017). Flourish or fail? the risky reward of elite high school admission in mexico city. *The Journal of Human Resources*, *52*(3), 756–799. Retrieved 2025-10-14, from http://www.jstor.org/stable/26450351

Fryer, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, *126*(4), 1755–1798.

Glewwe, P., Ilias, N., & Kremer, M. (2010, July). Teacher incentives. *American Economic Journal: Applied Economics*, *2*(3), 205–27. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.2.3.205 doi: 10.1257/app.2.3.205

Gneezy, U., List, J. A., Livingston, J. A., Sadoff, S., Qin, X., & Xu, Y. (2017, November). *Measuring success in education: The role of effort on the test itself* (Working Paper No. 24004). National Bureau of Economic Research. doi: 10.3386/w24004

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies*, *29*(1), 1–17.

Instituto Nacional Electoral. (2018). *Estadísticas y resultados electorales: Elección federal 2018.* https://www.ine.mx.

Instituto Nacional Electoral. (2021). *Estadísticas y resultados electorales: Elección federal 2021.* https://www.ine.mx.

Instituto Nacional para la Evaluación de la Educación. (2018). *Resultados de PLANEA* (Tech. Rep.). Retrieved from https://www.inee.edu.mx/evaluaciones/planea/resultados-planea/

Lavy, V. (2020, 01). Teachers' pay for performance in the long-run: The dynamic pattern of treatment effects on students' educational and labour market outcomes in adulthood. *The*

Review of Economic Studies, *87*(5), 2322-2355. Retrieved from https://doi.org/10.1093/restud/rdaa002 doi: 10.1093/restud/rdaa002

Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, *8*(4), 183–219.

Loyalka, P., Sylvia, S., Liu, C., Chu, J., & Shi, Y. (2019). Pay by design: Teacher performance pay design and the distribution of student achievement. *Journal of Labor Economics*, *37*(3), 621-662. Retrieved from https://doi.org/10.1086/702625 doi: 10.1086/702625

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. *The Quarterly Journal of Economics*, *134*(3), 1627-1673.

Mbiti, I., Romero, M., & Schipper, Y. (2023, 04). Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania. *The Economic Journal*.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy*, *119*(1), 39-77. Retrieved from https://doi.org/10.1086/659655 doi: 10.1086/659655

Santiago, P., McGregor, I., Nusche, D., Ravela, P., & Toledo, D. (2012). *Oecd reviews of evaluation and assessment in education: Mexico*. OECD publishing.

Todd, P., & Wolpin, K. I. (2018). Accounting for mathematics performance of high school students in mexico: Estimating a coordination game in the classroom. *Journal of Political Economy*, *126*(6), 2608–2650.

World Bank. (2017a). *Government expenditure on education, total (% of GDP) - Mexico.* (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS?locations=MX)

World Bank. (2017b). *Primary completion rate, total (% of relevant age group).* (data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.PRM.CMPT.ZS?locations=MX)

A Additional tables and figures

A.1 Tables

Table A.1: Balance across experimental groups in missing/incorrect CURPS

	Mean (SD) for each group				p-value
	Control	Students	Teacher	Both	(equality)
	(1)	(2)	(3)	(4)	(5)
Panel A: All cohorts					
Valid CURP (%)	0.99	0.99	0.99	0.99	0.605
	(0.10)	(0.11)	(0.12)	(0.11)	
	[30,992]	[24,256]	[21,976]	[20,618]	
Panel B: Cohort 1					
Valid CURP (%)	1.00	0.99	0.99	0.99	0.763
	(0.07)	(0.09)	(0.10)	(0.11)	
	[4,204]	[3,377]	[3,100]	[2,704]	
Panel C: Cohort 2					
Valid CURP (%)	0.99	0.99	0.98	0.98	0.756
	(0.11)	(0.10)	(0.13)	(0.13)	
	[5,295]	[4,193]	[3,836]	[3,530]	
Panel D: Cohort 3					
Valid CURP (%)	0.98	0.97	0.97	0.98	0.661
	(0.14)	(0.17)	(0.16)	(0.14)	
	[7,222]	[5,714]	[5,164]	[4,675]	
Panel E: Cohort 4					
Valid CURP (%)	0.99	0.99	0.99	0.99	0.489
	(0.08)	(0.09)	(0.11)	(0.09)	
	[7,169]	[5,507]	[4,982]	[4,814]	
Panel F: Cohort 5					
Valid CURP (%)	0.99	1.00	0.99	0.99	0.798
	(0.08)	(0.07)	(0.07)	(0.08)	
	[7,102]	[5,465]	[4,894]	[4,895]	

Notes: Columns 1–4 show the mean, standard deviation (in parentheses) and number of observations (in square brackets) for each of the following groups: Control, students exposed to student incentives, students exposed to teacher incentives, and students exposed to students and teacher incentives. Column 5 reports the p-value of whether the mean is the same across all groups. The equality test considers the randomization design (i.e., it includes strata fixed effects and clusters standard errors at the school level). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.2: Balance across experimental groups for the core sample

	Mean (SD) for each group				p-value
	Control	Students	Teacher	Both	(equality)
	(1)	(2)	(3)	(4)	(5)
Age (Jan 1, 2011)	17.90	17.83	18.04	17.78	0.127
	(10.41)	(1.68)	(17.29)	(1.57)	
	[30,775]	[24,062]	[21,830]	[20,462]	
Female	0.48	0.48	0.48	0.49	0.711
	(0.50)	(0.50)	(0.50)	(0.50)	
	[30,778]	[24,062]	[21,830]	[20,462]	
Missing ENLACE 9th grade score (Spanish or Math)	0.26	0.28	0.27	0.24	0.790
	(0.44)	(0.45)	(0.44)	(0.43)	
	[30,778]	[24,062]	[21,830]	[20,462]	
ENLACE 9th grade score (Spanish)	505.65	507.32	509.13	513.92	0.672
	(93.55)	(95.23)	(95.53)	(97.06)	
	[22,933]	[17,504]	[15,981]	[15,674]	
ENLACE 9th grade score (Math)	517.41	517.05	519.68	526.42	0.614
	(98.64)	(102.11)	(98.62)	(104.72)	
	[22,862]	[17,369]	[15,946]	[15,569]	

Notes: Columns 1–4 show the mean, standard deviation (in parentheses) and number of observations (in square brackets) for each of the following groups: Control, students exposed to student incentives, students exposed to teacher incentives, and students exposed to students and teacher incentives. Column 5 reports the p-value of whether the mean is the same across all groups. The equality test considers the randomization design (i.e., it includes strata fixed effects and clusters standard errors at the school level). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.3: Balance across experimental groups for the core sample by cohort

	Mean (SD) for each group				p-value
	Control	Students	Teacher	Both	(equality)
	(1)	(2)	(3)	(4)	(5)
Panel A: Cohort 1					
Age (Jan 1, 2011)	20.43	19.99	20.05	19.97	0.715
	(27.92)	(1.01)	(0.90)	(1.08)	
	[4,161]	[3,374]	[3,101]	[2,705]	
Female	0.52	0.53	0.51	0.53	0.765
	(0.50)	(0.50)	(0.50)	(0.50)	
	[4,162]	[3,374]	[3,101]	[2,705]	
Missing ENLACE 9th grade score (Spanish or Math)	0.59	0.64	0.66	0.53	0.357
	(0.49)	(0.48)	(0.47)	(0.50)	
	[4,162]	[3,374]	[3,101]	[2,705]	
ENLACE 9th grade score (Spanish)	513.26	505.29	512.97	506.90	0.886
	(91.89)	(86.48)	(93.50)	(95.05)	
	[1,716]	[1,229]	[1,051]	[1,285]	
ENLACE 9th grade score (Math)	514.32	499.05	518.48	513.36	0.226
	(96.14)	(92.77)	(95.58)	(98.20)	
	[1,719]	[1,231]	[1,051]	[1,281]	
Panel B: Cohort 2					
Age (Jan 1, 2011)	19.06	19.03	19.08	19.03	0.965
	(1.47)	(1.44)	(0.94)	(0.81)	
	[5,250]	[4,150]	[3,822]	[3,485]	
Female	0.51	0.50	0.50	0.51	0.970
	(0.50)	(0.50)	(0.50)	(0.50)	
	[5,250]	[4,150]	[3,822]	[3,485]	
Missing ENLACE 9th grade score (Spanish or Math)	0.38	0.45	0.35	0.34	0.602
	(0.49)	(0.50)	(0.48)	(0.47)	
	[5,250]	[4,150]	[3,822]	[3,485]	
ENLACE 9th grade score (Spanish)	520.81	516.06	521.39	521.42	0.861
	(92.10)	(93.07)	(93.79)	(98.33)	
	[3,248]	[2,292]	[2,486]	[2,320]	
ENLACE 9th grade score (Math)	519.71	507.02	520.75	523.32	0.343
	(95.53)	(94.76)	(94.81)	(103.02)	
	[3,243]	[2,284]	[2,481]	[2,312]	
Panel C: Cohort 3					
Age (Jan 1, 2011)	18.13	18.07	18.10	18.07	0.703
	(0.94)	(0.92)	(0.85)	(0.82)	
	[7,119]	[5,604]	[5,077]	[4,608]	
Female	0.48	0.47	0.48	0.49	0.289
	(0.50)	(0.50)	(0.50)	(0.50)	
	[7,121]	[5,604]	[5,077]	[4,608]	
Missing ENLACE 9th grade score (Spanish or Math)	0.18	0.19	0.18	0.19	0.821
	(0.38)	(0.39)	(0.38)	(0.39)	
	[7,121]	[5,604]	[5,077]	[4,608]	
ENLACE 9th grade score (Spanish)	505.33	507.45	506.05	513.32	0.824
	(92.43)	(94.51)	(91.68)	(97.00)	
	[5,856]	[4,597]	[4,190]	[3,764]	
ENLACE 9th grade score (Math)	515.47	515.36	513.89	520.15	0.866
	(100.29)	(103.30)	(96.96)	(104.37)	
	[5,839]	[4,538]	[4,187]	[3,732]	
Panel D: Cohort 4					
Age (Jan 1, 2011)	17.12	17.07	17.49	17.08	0.641
	(0.98)	(0.95)	(25.62)	(0.82)	
	[7,128]	[5,478]	[4,941]	[4,779]	
Female	0.46	0.47	0.48	0.49	0.219
	(0.50)	(0.50)	(0.50)	(0.50)	
	[7,128]	[5,478]	[4,941]	[4,779]	
Missing ENLACE 9th grade score (Spanish or Math)	0.17	0.17	0.19	0.17	0.971
	(0.38)	(0.38)	(0.39)	(0.38)	
	[7,128]	[5,478]	[4,941]	[4,779]	
ENLACE 9th grade score (Spanish)	504.61	512.07	512.80	510.74	0.796
	(92.04)	(95.83)	(95.33)	(94.69)	
	[5,924]	[4,563]	[4,033]	[3,978]	
ENLACE 9th grade score (Math)	516.26	523.39	520.95	521.75	0.902
	(93.58)	(99.28)	(96.61)	(103.45)	
	[5,911]	[4,544]	[4,027]	[3,969]	
Panel E: Cohort 5					
Age (Jan 1, 2011)	16.11	16.08	16.47	16.07	0.697
	(1.04)	(0.90)	(25.75)	(0.81)	
	[7,117]	[5,456]	[4,889]	[4,885]	
Female	0.48	0.46	0.46	0.47	0.744
	(0.50)	(0.50)	(0.50)	(0.50)	
	[7,117]	[5,456]	[4,889]	[4,885]	
Missing ENLACE 9th grade score (Spanish or Math)	0.14	0.13	0.14	0.12	0.899
	(0.34)	(0.33)	(0.35)	(0.33)	
	[7,117]	[5,456]	[4,889]	[4,885]	
ENLACE 9th grade score (Spanish)	496.87	499.08	500.52	515.43	0.058*
	(96.08)	(97.82)	(99.92)	(98.91)	
	[6,189]	[4,823]	[4,221]	[4,327]	
ENLACE 9th grade score (Math)	519.99	522.08	523.89	541.81	0.059*
	(103.92)	(108.21)	(104.73)	(107.35)	
	[6,150]	[4,772]	[4,200]	[4,275]	

Notes: Columns 1–4 show the mean, standard deviation (in parentheses) and number of observations (in square brackets) for each of the following groups: Control, students exposed to student incentives, students exposed to teacher incentives, and students exposed to students and teacher incentives. Column 5 reports the p-value of whether the mean is the same across all groups. The equality test considers the randomization design (i.e., it includes strata fixed effects and clusters standard errors at the school level). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4: Heterogeneity on the likelihood of taking the ENLACE 12 exam (intent-to-treat)

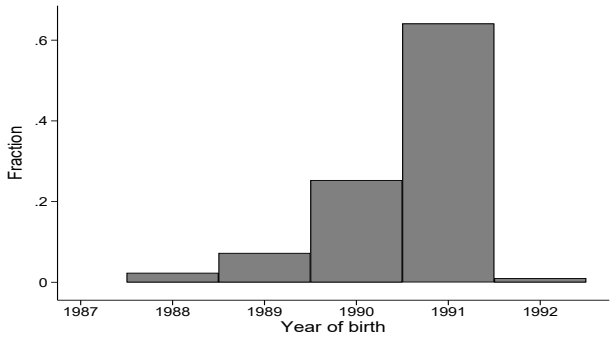
	Took Grade 12 ENLACE					
	(1)	(2)	(3)	(4)	(5)	(6)
Students	-2.33	0.94	-2.99	0.19	-4.33	-2.35
	(3.42)	(2.48)	(3.70)	(2.49)	(3.09)	(2.07)
Teachers	-1.30	-0.12	-1.65	-0.68	-2.80	-2.10
	(2.53)	(2.12)	(2.60)	(2.15)	(2.47)	(2.07)
Both	-5.51***	-3.88*	-6.06***	-4.39**	-5.03**	-4.54**
	(2.04)	(2.03)	(1.93)	(1.99)	(1.98)	(1.98)
Students × Grade 9 Spanish score above median	-2.77	-4.29*				
	(2.36)	(2.47)				
Teachers × Grade 9 Spanish score above median	-0.18	-1.29				
	(1.60)	(1.60)				
Both × Grade 9 Spanish score above median	0.81	-0.51				
	(2.05)	(1.96)				
Students × Grade 9 math score above median			-1.04	-2.86		
			(1.79)	(2.13)		
Teachers × Grade 9 math score above median			0.71	-0.14		
			(1.60)	(1.63)		
Both × Grade 9 math score above median			1.93	0.56		
			(2.37)	(2.22)		
Students × Female					0.81	1.06
					(1.19)	(1.16)
Teachers × Female					1.60	1.39
					(1.18)	(1.15)
Both × Female					1.79	1.92
					(1.36)	(1.36)
Control mean	.49	.49	.48	.48	.48	.48
p-value equality	.14	.05	.14	.073	.69	.46
R^2	.078	.1	.08	.1	.072	.11
N. of obs.	72,317	72,317	72,317	72,317	97,176	97,176
Controls	No	Yes	No	Yes	No	Yes

Notes: All specifications include strata, cohort, gender, and birth-year fixed effects. The dependent variable is whether a student ever takes the ENLACE Grade 12 exam (0/100). Columns 1–2 interact treatment indicators with an indicator for Grade 9 Spanish score above the median; Columns 3–4 interact with Grade 9 math score above the median; Columns 5–6 interact with an indicator for female. Odd-numbered columns include no additional controls beyond the fixed effects listed above. Even-numbered columns add: school average Grade 12 ENLACE scores in 2008 (when available), student-level Grade 9 ENLACE scores (when available), an indicator for schools with a new principal in 2008–2009, distance to the nearest city, the historical Grade 10-to-graduation rate, average students per classroom, share of teachers with tertiary degrees, and total school enrollment. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

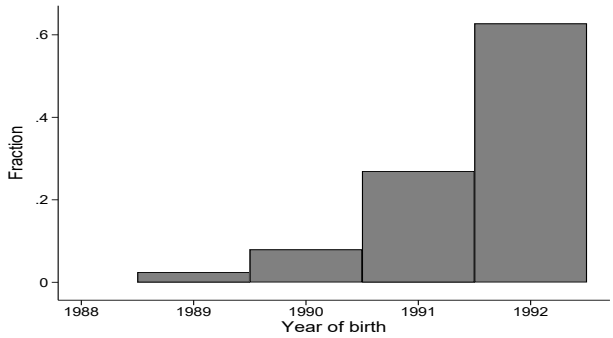
A.2 Figures

Figure A.1: Distribution of birth years for different cohorts

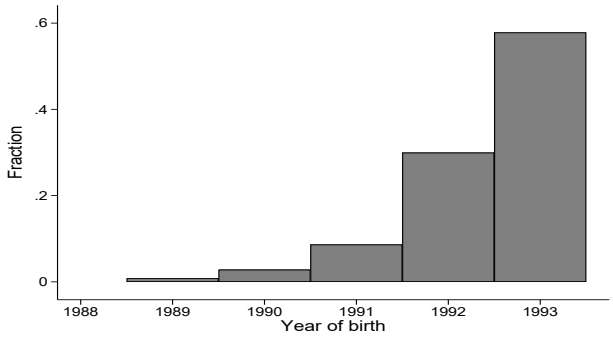
(a) Cohort 1/class of 2009



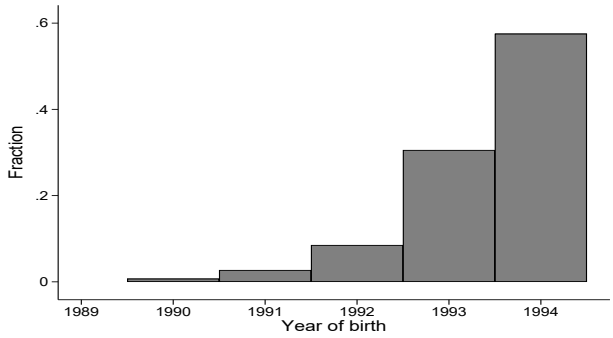
(b) Cohort 2/class of 2010



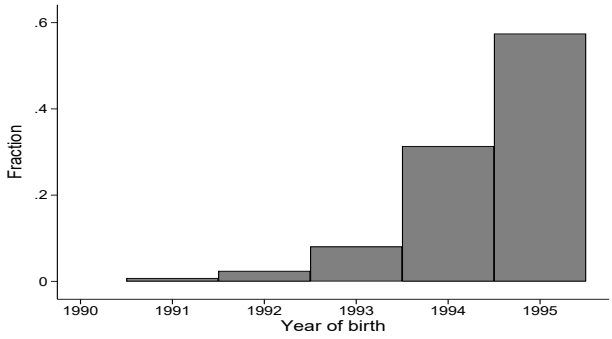
(c) Cohort 3/class of 2011



(d) Cohort 4/class of 2012



(e) Cohort 5/class of 2013



Note: These figures plot the fraction of students born in different years for different cohorts. Students born before 1980 or after 1997 (inclusive) are excluded (∼0.09% of the data). For each cohort, the bottom 1% and top 1% of the data are trimmed before plotting the histogram.