Acing the Test: Educational Effects of the *SaberEs* Test Preparation Program in Colombia

Christian Posso [1] Estefanía Saravia [2] **Pablo Uribe** [3]

[1]Banco de la República

[2]Icfes & UCLA

[3]**EAFIT - World Bank**

IEA World Congress - December 13, 2023

Contents

Introduction

- Dramatic expansion in post-primary education in low- and middle-income countries (Ferreyra, Avitabile, Paz, Botero, & Urzúa, 2017; World Bank, 2018).

- There are important gaps in achievement and quality of education as measured by test scores (Angrist & Lavy, 2009; Gneezy et al., 2019).

- Standardized tests affect the transition to higher education and labor market outcomes (Bond, Bulman, Li, & Smith, 2018; Brunello & Kiss, 2022).

- Scores determine eligibility for financial aid (Bernal & Penney, 2019; Bruce & Carruthers, 2014; Gurantz & Odle, 2020; Londoño-Vélez, Rodríguez, & Sánchez, 2020; Melguizo, Sanchez, & Velasco, 2016).

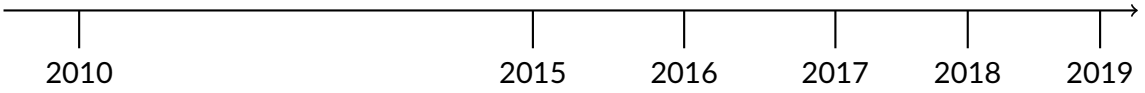
SaberEs program

- Secretary of Education of Medellin $\longrightarrow$ 2016.
- 2016-2019 Development Plan.
- Develop skills that strengthen preparation for standardized tests like Saber 11.
- Additional installed capacity and vocational guidance components (applied after Saber 11 test).
- Low-cost program: (dollars 2015) 1.4 USD million in total, 32.6 USD per student.
- 2 companies hired:
  - Separate set of schools.
  - Grades 8 to 11.
  - Teacher training $\longrightarrow$ Trained schools' principals and coordinators $\longrightarrow$ Teachers train students during school hours $\longrightarrow$ Simulation tests $\longrightarrow$ Feedback sessions.

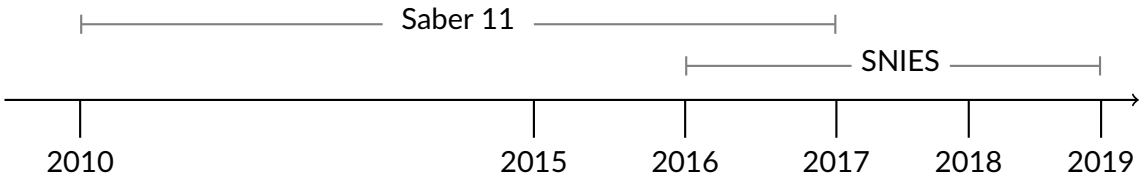
Our paper

- Provide new evidence on the effectiveness of standardized test preparation programs.
- Especially important since most of this evidence suffers from self-selection bias.
- Three main questions:
  1. Does the *SaberEs* program affect student learning gains measured by *Saber 11* scores?
  2. Does the program affect access to tertiary education?
  3. What mechanisms made *SaberEs* successful in increasing access to tertiary education programs?

Timeline

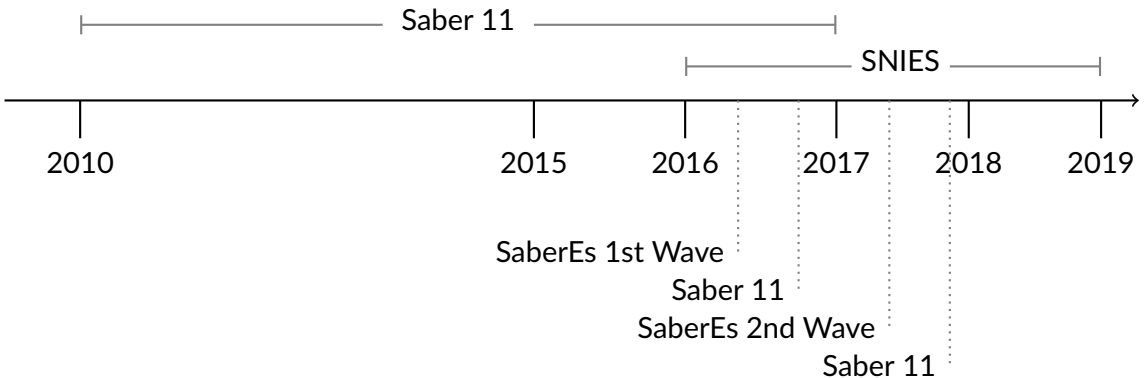


Timeline



Other datasets: ICETEX (2018-2019), Sapiencia (2016-2019), Saber TyT (2016-2019), Ser Pilo Paga (2015-2016), Olimpiadas del conocimiento (2015-2016).

Timeline


Saber 11
SNIES
2010
2015
2016
2017
2018
2019
SaberEs 1st Wave
Saber 11
SaberEs 2nd Wave
Saber 11

Contents

Data

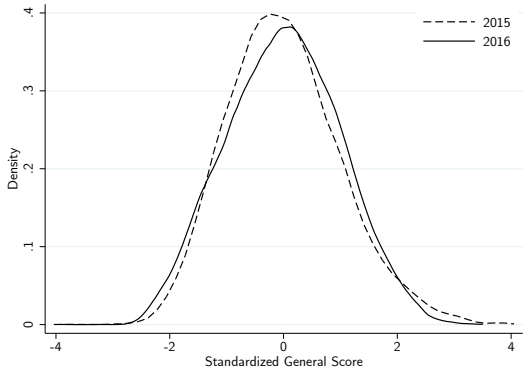
- Administrative data from eight main sources.
- Saber 11 had a structural change in 2014 $\longrightarrow$ Student's rank as variable of interest (0-100) to ensure comparability, following Laajaj, Moya, and Sánchez (2022).

$$Rank_i = \frac{x_i - min(x)}{max(x) - min(x)} * 100$$

- 2x2 (2015-2016) and dynamic difference-in-differences specifications (2010-2017).

Data

- Distribution of the standardized scores in 2015-2016.
- Right shift concentrated along the median students.



Summary Statistics

	Mean	SD	Min	Max
Panel A: Test Scores				
General	258.69	42.14	13	450
Reading	52.85	8.93	0	100
Math	51.13	10.52	0	100
Science	51.48	9.11	0	100
Social Studies	51.52	10.06	0	93
English	51.66	10.37	0	100
Panel B: Higher Education and Financial Aid				
Access to higher education	0.60	0.49	0	1
Access to short-cycle	0.31	0.46	0	1
Access to university	0.33	0.47	0	1
Access to STEM	0.26	0.44	0	1
Access to professional STEM	0.16	0.37	0	1
Access to short-cycle STEM	0.13	0.34	0	1
Received financial aid	0.05	0.22	0	1
Received *Ser Pilo Paga*	0.03	0.16	0	1
Panel C: Treatment				
Treated	0.66	0.47	0	1
Treated Tres Editores	0.46	0.50	0	1
Treated Avancemos	0.20	0.40	0	1

	Mean	SD	Min	Max
Panel D: Covariates				
Female	0.57	0.50	0	1
TV	0.80	0.40	0	1
Oven	0.60	0.49	0	1
Landline	0.85	0.36	0	1
Microwave	0.50	0.50	0	1
PC	0.78	0.42	0	1
Car	0.16	0.37	0	1
Internet	0.77	0.42	0	1
Washing machine	0.81	0.39	0	1
DVD	0.61	0.49	0	1
NSE 1	0.03	0.18	0	1
NSE 2	0.33	0.47	0	1
NSE 3	0.62	0.48	0	1
NSE 4	0.02	0.13	0	1
Employed	0.06	0.23	0	1
Parent's education	0.10	0.30	0	1
High income	0.07	0.26	0	1
High stratum	0.04	0.20	0	1
Household floor	0.42	0.49	0	1
> 6 People in household	0.20	0.40	0	1
> 3 Rooms in household	0.61	0.49	0	1

Contents

Empirical Strategy - 2x2

We estimate a simple difference-in-differences regression as:

$$Y_{ict} = \alpha + \beta_0 \text{Treated}_c + \beta_1 \text{Post}_t + \beta_2 \text{Treated}^* \text{Post}_{ct} + X'_{ict}\delta + \varepsilon_{ict}$$

where:

- Y_{ict} is the general rank of student i from school c in period t.
- $Treated_c$ is a dummy variable indicating whether school c is treated.
- $Post_t$ takes a value of 1 if the student's test application year is 2016.
- $Treated * Post_{ct}$ is their interaction.
- X'_{ict} is a vector of controls.
- ε_{ict} is the error term.

Empirical Strategy - 2x2

We also estimate a two-way fixed effects regression as:

$$Y_{ict} = \alpha + \theta_1 \text{Treated*Post}_{ct} + \psi_c + \gamma_t + \mu_{ict}$$

where ψ_c and γ_t are the school and year fixed effects, respectively. μ_{ict} is the error term.

Empirical Strategy - 2x2

We use alternatives that better handle the inclusion of covariates:

- Outcome regression (Heckman, Ichimura, & Todd, 1997).
- Hájek (1971) type inverse probability weighting (IPW) with normalized weights.
- Sant'Anna and Zhao (2020) doubly robust improved difference-in-differences estimator for repeated cross sections.
- RIF regressions (Firpo, Fortin, & Lemieux, 2009) $\longrightarrow$ Effects on the unconditional quantiles.

Empirical Strategy - Dynamic

- Observations from 2010 to 2016 (2017) $\longrightarrow$ Multiple time periods and one (two) year (years) of treatment.

- When treatment is staggered TWFE would potentially be biased due to the presence of heterogeneous effects (Borusyak & Jaravel, 2017; De Chaisemartin & d'Haultfoeuille, 2020).

- Estimator is a weighted average of all 2x2 comparisons and includes "forbidden comparisons" (Goodman-Bacon, 2021).

Empirical Strategy - Dynamic (non-staggered)

We estimate an event studies regression as:

$$Y_{ict} = \alpha_0 + \sum_{h=2010}^{2016} \beta_h [t * \text{Treated}_{ch} = h] + \psi_c + \gamma_t + u_{ict} \quad \forall \quad h \neq 2015$$

where Y_{ict} is the outcome of student i from school c at time t, and $t * Treated_{ct}$ are the interactions of year and treatment status for each of the leads or lags (h). ψ_c and γ_t are the school and year fixed effects respectively, and u_{ict} is the error term.

Empirical Strategy - Dynamic (staggered)

- We use the Callaway and Sant'Anna (2021) estimator.
  - Simple aggregation.
  - Event study aggregation.

- Alternative specification proposed by Borusyak, Jaravel, and Spiess (2021).
  - Stronger assumption about parallel trends could lead to a larger bias (Roth, Sant'Anna, Bilinski, & Poe, 2022).

- Calculate possible bias from pre-testing (Roth, 2022) and conduct a sensitivity analysis (Rambachan & Roth, 2023).

Contents

Results - 2x2

- Statistically significant and positive effect on the average student's rank.
- Holds to more robust specifications like the doubly robust one.
- Using the standardized test scores further proves the robustness of the results.

▸ See Results

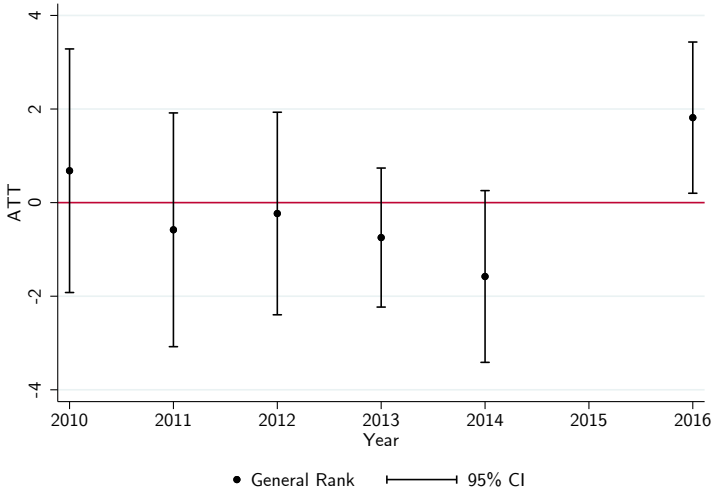
	(1) DiD	(2) DiD	(3) TWFE	(4) OR	(5) IPW	(6) DR
SaberEs effect (β)	2.965***	2.559***	1.511**	2.222**	2.693***	2.233**
	(0.976)	(0.886)	(0.741)	(0.917)	(1.032)	(0.916)
Observations	35,495	35,484	35,484	35,484	35,484	35,484
Controls	NO	YES	NO	YES	YES	YES

Results - 2x2

- Statistically significant and positive effect on the average student's rank.
- Holds to more robust specifications like the doubly robust one.
- Using the standardized test scores further proves the robustness of the results.
  ▸ See Results
- 22.9% reduction in the rank's gap between treated and untreated students.
- Mainly driven by math improvements
  ▸ See Results

	(1) DiD	(2) DiD	(3) TWFE	(4) OR	(5) IPW	(6) DR
SaberEs effect (β)	2.965***	2.559***	1.511**	2.222**	2.693***	2.233**
	(0.976)	(0.886)	(0.741)	(0.917)	(1.032)	(0.916)
Gap reduction	30.6%	26.4%	15.6%	22.9%	27.8%	22.9%
Observations	35,495	35,484	35,484	35,484	35,484	35,484
Controls	NO	YES	NO	YES	YES	YES

Results - Dynamic (non-staggered)



Results - Dynamic (staggered)

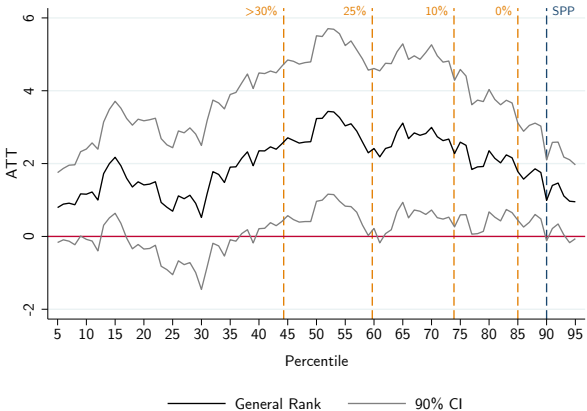
- Extending the sample from 2010 to 2017 yields better results.
- Effects are robust to both specifications.
- The program reduced the gap by around 30-40%.
- Results are robust when using the standardized test scores. ▸ See Results
- The event study aggregations with balanced groups show a similar pattern. ▸ See Results
- Power and sensitivity analyses on the pre-trends further show the robustness of the results.
  ▸ Power analysis ▸ Sensitivity analysis

	(1) C&S	(2) BJS	(3) C&S	(4) BJS
SaberEs effect (β)	3.715***	2.711***	3.598***	2.688***
	(0.785)	(0.497)	(0.820)	(0.462)
Gap reduction	38.3%	28.0%	37.1%	27.7%
Observations	147,656	147,554	147,656	70,859
Controls	NO	NO	YES	YES

Notes: Standard errors clustered at the school level. C&S relates to the "simple" aggregation from Callaway and Sant'Anna (2021). BJS relates to the estimator proposed by Borusyak et al. (2021). Controls include gender, household goods and services (computer, car, internet and washing machine), parents education, and stratum. *p<.05; **p<.01; ***p<.001

Results - Heterogeneous effects on the outcome distribution

- Effects are evidenced above the 40th percentile of the students' rank distribution.
- They are similar when looking at the effects on the distribution of standardized test scores. ▸ See Figure
- There are positive effects on students above the SPP cut-off.



Results - Higher education

	Higher Education			Short-cycle			Professional		
	(1) 1 year	(2) 2 years	(3) 3 years	(4) 1 year	(5) 2 years	(6) 3 years	(7) 1 year	(8) 2 years	(9) 3 years
SaberEs effect (β)	0.033**	0.039***	0.024**	0.026**	0.023**	0.010	0.006	0.015	0.014
	(0.015)	(0.015)	(0.012)	(0.013)	(0.011)	(0.011)	(0.011)	(0.012)	(0.010)
Observations	35,484	35,484	35,484	35,484	35,484	35,484	35,484	35,484	35,484
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean Control 2015	0.511	0.559	0.587	0.219	0.230	0.224	0.292	0.329	0.361

Notes: Standard errors clustered at the school level. Results come from a doubly robust estimation as in Sant'Anna and Zhao (2020). The columns indicate access to each outcome 1, 2 and 3 years after students graduate from high school. *p<.05; **p<.01; ***p<.001

Results - Higher Education

	(1) STEM	(2) Professional STEM	(3) Short-cycle STEM
SaberEs effect (β)	0.021*	0.011	0.019**
	(0.011)	(0.009)	(0.008)
Observations	35,484	35,484	35,484
Controls	YES	YES	YES
Mean Control 2015	0.287	0.173	0.152

Notes: Standard errors clustered at the school level. Results come from a doubly robust estimation as in Sant'Anna and Zhao (2020). *p<.05; **p<.01; ***p<.001

Results - Graduation from short-cycle education

Graduation from short-cycle education increases for all programs, and even for STEM programs.

	(1) All short-cycle programs	(2) Short-cycle STEM programs
SaberEs effect (β)	0.023**	0.010*
	(0.010)	(0.006)
Observations	35,484	35,484
Controls	YES	YES
Mean Control 2015	0.152	0.062

Notes: Standard errors clustered at the school level. Results come from a doubly robust estimation as in Sant'Anna and Zhao (2020). *p<.05; **p<.01; ***p<.001

Potential Mechanisms

For the majority of the population:

- Accumulation of specific human capital $\longrightarrow$ Access to universities that require admission exams (UdeA, Nacional & SENA). ✓ See results
- Motivational effect (effects on *Olimpiadas del conocimiento*). × See results
- Access to financial aid (ICETEX & Sapiencia). × See results

For elite students:

- Access to Ser Pilo Paga. ✓ See results

Contents

Conclusion

- One of the few papers to analyze these types of policies for socioeconomically disadvantaged students in Latin America aside from Gómez, Bernal, and Herrera (2020).

- We take advantage of granular administrative data to identify the causal effect of SaberEs on students' academic performance.

- We use state of the art econometric methods in a difference-in-differences estimation.

Conclusion

- One of the few papers to analyze these types of policies for socioeconomically disadvantaged students in Latin America aside from Gómez et al. (2020).
- We take advantage of granular administrative data to identify the causal effect of SaberEs on students' academic performance.
- We use state of the art econometric methods in a difference-in-differences estimation.
- We find a positive effect of over 2 points on the average student's rank in the test $\longrightarrow$ 22.9% reduction in the pre-existing gap.
- In terms of higher education, we find a positive effect on access to short-cycle and STEM programs. Also, the program positively affected graduation from short-cycle programs.
- A limitation of our paper is the absence of a cost-benefit analysis. However, we expect the net present value of benefits to be large and positive.

References I

Angrist, J., & Lavy, V. (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American economic review*, *99*(4), 1384–1414.

Bernal, G. L., & Penney, J. (2019). Scholarships and student effort: Evidence from colombia's ser pilo paga program. *Economics of Education Review*, *72*, 121–130.

Bond, T. N., Bulman, G., Li, X., & Smith, J. (2018). Updating human capital decisions: Evidence from sat score shocks and college applications. *Journal of Labor Economics*, *36*(3), 807–839.

Borusyak, K., & Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.

Borusyak, K., Jaravel, X., & Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Bruce, D. J., & Carruthers, C. K. (2014). Jackpot? the impact of lottery scholarships on enrollment in tennessee. *Journal of Urban Economics*, *81*, 30–44.

Brunello, G., & Kiss, D. (2022). Math scores in high stakes grades. *Economics of Education Review*, *87*, 102219.

References II

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–96.

Ferreyra, M. M. (2021). Landscape of short-cycle programs in latin america and the caribbean. *The Fast Track to New Skills*, 33.

Ferreyra, M. M., Avitabile, C., Paz, F. H., Botero, J., & Urzúa, S. (2017). *At a crossroads: higher education in latin america and the caribbean*. World Bank Publications.

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: the role of effort on the test itself. *American Economic Review: Insights*, *1*(3), 291–308.

References III

Gómez, S. C., Bernal, G. L., & Herrera, P. (2020). Test preparation and students' performance: The case of the colombian high school exit exam. *Cuadernos de Economía*, *39*(79), 31–72.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Gurantz, O., & Odle, T. K. (2020). The impact of merit aid on college choice and degree attainment: Reexamining florida's bright futures program. *Educational Evaluation and Policy Analysis*, 01623737211030489.

Hájek, J. (1971). Discussion of 'an essay on the logical foundations of survey sampling, part i', by d. basu. *Foundations of statistical inference*, 326.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, *64*(4), 605–654.

References IV

Laajaj, R., Moya, A., & Sánchez, F. (2022). Equality of opportunity and human capital accumulation: Motivational effect of a nationwide scholarship in colombia. *Journal of Development Economics*, *154*, 102754.

Londoño-Vélez, J., Rodríguez, C., & Sánchez, F. (2020). Upstream and downstream impacts of college merit-based financial aid for low-income students: Ser pilo paga in colombia. *American Economic Journal: Economic Policy*, *12*(2), 193–227.

Melguizo, T., Sanchez, F., & Velasco, T. (2016). Credit for low-income students and access to and academic performance in higher education in colombia: A regression discontinuity approach. *World development*, *80*, 61–77.

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*.

Roth, J. (2022). Pre-test with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2022). *What's trending in difference-in-differences? a synthesis of the recent econometrics literature.*

References V

Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101–122.

World Bank. (2018). *World bank education overview : Higher education (english)*. World Bank Group.

Contents

Saber 11

- High school exit exam administered by ICFES.
- Compulsory test with compliance rates above 90% (Bernal & Penney, 2019).
- 500,000 students take it every year (March-August).
- Structural change in 2014 $\longrightarrow$ 5 subject areas with scores between 0-100.
- ICFES offered a familiarization test that costed $30 USD (Bernal & Penney, 2019).
- Private companies offer courses (mostly used by private schools).

Higher education

- Public and private institutions with admission processes each semester.
- Saber 11 plays a central role in the admission processes (Londoño-Vélez et al., 2020).
- Higher education costs in Colombia are relatively high (Ferreyra, 2021).
- Biggest public universities have highly competitive admission processes $\longrightarrow$ Highest-achieving students enroll.
- Enrollment in private institutions for low-income students is mainly driven by funding.

Results - 2x2

Table: Main results: Standardized score.

	(1) DiD	(2) DiD	(3) TWFE	(4) OR	(5) IPW	(6) DR
SaberEs effect (β)	0.104***	0.089***	0.053**	0.073**	0.092**	0.074**
	(0.034)	(0.030)	(0.026)	(0.032)	(0.036)	(0.032)
Gap reduction	31.1%	26.6%	15.8%	22.0%	27.4%	22.3%
Observations	35,495	35,484	35,484	35,484	35,484	35,484
Controls	NO	YES	YES	YES	YES	YES

Back

Results - 2x2 specific

Figure: Rank.

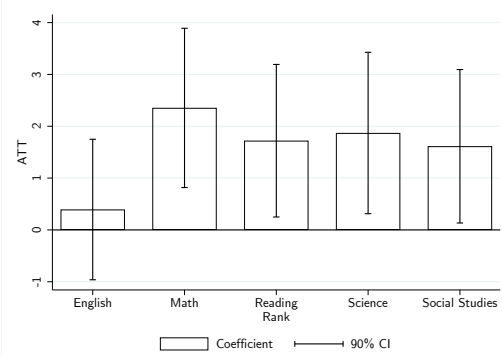
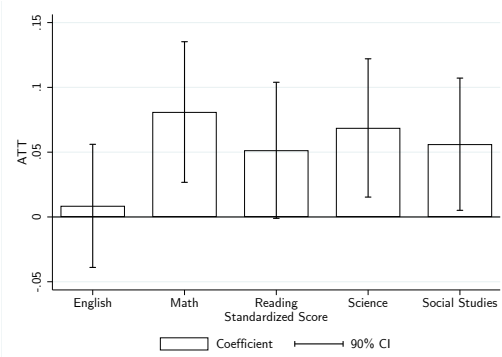


Figure: Standardized scores.



Back

Results - 2x2

Back

Results - Dynamic

Table: Dynamic Results: Standardized Score.

	(1) C&S	(2) BJS	(3) C&S	(4) BJS
SaberEs effect (β)	0.131***	0.099***	0.123***	0.094***
	(0.028)	(0.016)	(0.029)	(0.015)
Gap reduction	39.2%	29.6%	36.8%	28.1%
Observations	147,656	147,554	147,656	70,859
Controls	NO	NO	YES	YES

Notes: Standard errors clustered at the school level. C&S relates to the "simple" aggregation from Callaway and Sant'Anna (2021). BJS relates to the estimator proposed by Borusyak et al. (2021). Controls include gender, household goods and services (computer, car, internet and washing machine), parents education, and stratum. *p<.05; **p<.01; ***p<.001

▸ Back

Results - Dynamic

Figure: Average Effect on Student's Rank by Length of Exposure.

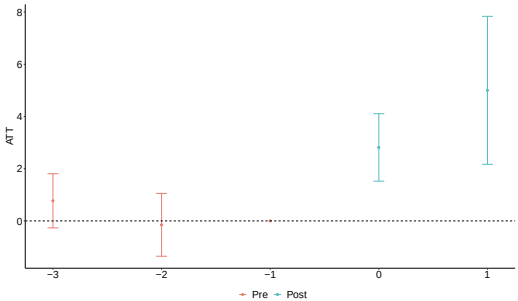
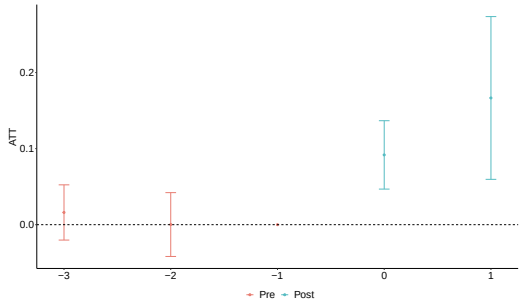


Figure: Average Effect on Student's Standardized Scores by Length of Exposure.



Back

Power Analysis

Table: Power analysis: bias from hypothesized trend

	(1) Estimate	(2) Slope	(3) Likelihood ratio
General Rank	3.715	0.462	0.009
Standardized General Score	0.131	0.016	0.009

Notes: Column 1 displays the estimated "simple" coefficient from 21 and 2. Column 2 shows the pre-trend that has 50% power of being detected (hypothesized trend). Column 3 shows the likelihood ratio.

Back

Sensitivity Analysis

Figure: Sensitivity analysis: general rank.

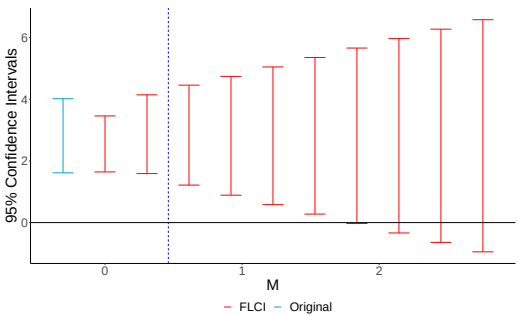
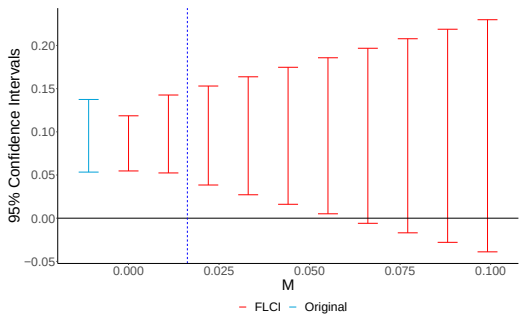


Figure: Sensitivity analysis: standardized general score.



Back

Specific human capital mechanism

Table: Effects on access to short-cycle programs in institutions with *Saber 11*-like admission exams

	Admission exam			No admission exam		
	(1) 1 year	(2) 2 years	(3) 3 years	(4) 1 year	(5) 2 years	(6) 3 years
SaberEs effect (β)	0.026**	0.024**	0.010	0.000	-0.001	0.000
	(0.012)	(0.010)	(0.010)	(0.007)	(0.006)	(0.008)
Observations	35,484	35,484	35,484	35,484	35,484	35,484
Controls	YES	YES	YES	YES	YES	YES
Mean Control 2015	0.132	0.141	0.126	0.0865	0.0896	0.0982

Notes: Standard errors clustered at the school level. Results come from a doubly robust estimation as in Sant'Anna and Zhao (2020). Institutions with *Saber 11*-like admission exams that offer short-cycle programs are SENA and UdeA. *p<.05; **p<.01; ***p<.001

▸ Back

Motivational effect mechanism

Table: Effects on student's rank in *Olimpiadas del Conocimiento*

	(1) Grade 10	(2) Grade 11	(3) Joint
SaberEs effect (β)	-0.271	-1.647	-0.992
	(1.056)	(1.213)	(0.901)
Observations	35,852	31,592	67,444
Controls	YES	YES	YES

Notes: Standard errors clustered at the school level. Results come from a doubly robust estimation as in Sant'Anna and Zhao (2020). All specifications control for stratum, family income, parent's education, mobile phone ownership and student's working status. *p<.05; **p<.01; ***p<.001

▸ Back

Financial aid mechanism

Table: Effects on access to financial aid and *Ser Pilo Paga*

	(1) Higher Education	(2) Financial Aid	(3) Ser Pilo Paga
SaberEs effect (β)	0.037***	0.005	0.010***
	(0.013)	(0.005)	(0.004)
Observations	35,484	35,484	35,484
Controls	YES	YES	YES
Mean Control 2015	0.677	0.0530	0.0464

Notes: Standard errors clustered at the school level. Results come from a doubly robust estimation as in Sant'Anna and Zhao (2020). *p<.05; **p<.01; ***p<.001

▸ Back

Merit-based scholarship mechanism

Table: Effects on access to SPP by type of program

	(1) Short-cycle SPP	(2) Professional SPP
SaberEs effect (β)	0.001*	0.010**
	(0.000)	(0.004)
Observations	35,484	35,484
Controls	YES	YES
Mean Control 2015	0.0007	0.0457

Notes: Standard errors clustered at the school level. Results come from a doubly robust estimation as in Sant'Anna and Zhao (2020). *p<.05; **p<.01; ***p<.001

Back