Evaluating the effectiveness of the Full-Time School Program in Mexico

Pablo Velazquez-Villa[1] Diego A. Luna-Bazaldua[2]
[1]Mexicanos Primero [2]Universidad Nacional Autónoma de México

Introduction

Intervention background and Research Question

The federal government of Mexico introduced an educational policy titled Full-Time School (ETC; *Escuela de Tiempo Completo* in Spanish) in 2008, which increased the number of hours in the classroom from 800 to 1,200 hours during the school year. The rationale behind the program was that more hours devoted to school activities would increase learning and other educational outcomes. From the beginning of its implementation, the selection of schools to be included in the program was not random. Thus, there have been difficulties in determining a causal impact of the intervention on educational achievement. In addition, over the years, the implementation of the program has increased nationwide. For instance, 2,012 schools were part of the ETC program in 2009, while this number increased to 24,507 by 2015.

Previous research has reported descriptive information indicating that after schools take part in this intervention, there is a decrease in the proportion of their students classified in lowest tier level of achievement used for the Mexican national educational test (ENLACE; *Evaluación Nacional de Logros Académicos en Centros Escolares*), while also increasing the proportion of students in the highest tier level of achievement (Gómez-Zermeño, Flores-Fahara, & Alemán de la Garza, 2013). Nevertheless, it is difficult to establish a causal link between the ETC program and the educational achievement of students and schools. Based on this background, the objective of this research was to determine whether ETC has had any impact on educational achievement as measured by ENLACE using analytical approaches proposed under the causal inference framework (Shadish, Cook, & Campbell, 2002).

Methods

Setting, population, research design

Because of the type and amount of longitudinal data available, this study combined the use of a difference-in-differences design with propensity score matching (Somers, Zhu, Jacob, & Bloom, 2013; Stuart et al., 2014). School-level educational achievement data on ENLACE was publicly available for all elementary public and private schools nationwide in Mexico from the school years 2006-07 – which was used as a pre-intervention baseline – up to 2013-14. This longitudinal data set also included information on whether the schools implemented ETC or not during any given school year, which permits the use of a difference-in-differences design (Somers et al., 2013) to compare schools that implemented the program and those who did not.

A total of 129 additional sociodemographic covariates about the local area where the schools were located were available from public data from the Mexican national census bureau (INEGI; *Instituto Nacional de Estadística y Geografía*). The information in the data set revealed that ETC had been implemented only in public and indigenous schools in Mexico, so all private or community schools were left out of the analysis. A comparison of the sociodemographic profile of schools that implemented ETC for at least one year and the remaining public schools

that had never implemented the program revealed remarkable differences on almost every sociodemographic variable as measured using covariate balance measures (i.e., absolute covariate bias). Thus, propensity scores were estimated using only sociodemographic covariates with little or no multicollinearity among them as the predictors of the intervention status. The propensity scores were computed using a generalized boosted model (McCaffrey et al., 2013; Ridgeway, McCaffrey, Morral, Griffin, & Burgette, 2016). Then, the propensity scores were used to perform a one-to-one match between intervention and comparison schools to reduce covariate bias in the matched sample (Austin, 2011; Caliendo & Kopeinig, 2008; Guo & Fraser, 2010). This process resulted in a mean absolute covariate bias reduction from 0.129 in the complete sample of public schools (N = 62,246) with sociodemographic information to 0.042 in the matched sample (N = 21,400).

Because of the presence of missing values in the outcomes, multiple imputation techniques were used to generate ten independent data sets with imputed values in the complete sample (Zachrisson, Dearing, Lekhal y Toppelberg, 2013; Van Buuren, 2007). The model described next was estimated with the complete sample, the matched sample, and matched samples with imputed values.

Hierarchical longitudinal model

Once the analytical sample was consolidated, a hierarchical model was estimated (Raudenbush & Bryk, 2002). Level one included the school-level scores in ENLACE Spanish and mathematics subtests of every school, as well as a dichotomous indicator to identify whether they had implemented ETC ($ETC_{ti} = 1$) or not ($ETC_{ti} = 0$) during a given school year. Level 2 included school level information and the sociodemographic covariates. Level 2 also included a dichotomous indicator to denote whether the school ever implemented ETC ($T_i = 1$) or not ($T_i = 0$). This multilevel structure in the data set allowed the estimation of the fixed-effects difference-in-difference model:

Level 1 $Score_{ti} = \pi_{0i} + \pi_{1i} * (ETC_{ti}) + e_{ti}$

Level 2 $\pi_{0i} = \beta_{00} + \boldsymbol{\beta}_{\cdot} * (\mathbf{covariates_i}) + r_{0i}$

$\pi_{1i} = \beta_{10} + \beta_{11} * (T_i)$

$\boldsymbol{\pi}_{\cdot \mathbf{i}} = \boldsymbol{\beta}_{\cdot \mathbf{0}}$

where the coefficient of interest is β_{11} (γ_{11} in the complete model) since it denotes the fixed-effect for the ETC program during those school years in which the program was actually implemented in the schools taking part of the intervention.

Results

Descriptive results indicate that intervention schools, on average, increased their test scores during the years that they implemented the program. The test score increases in Spanish and mathematics were higher than those of the comparison schools (see Table 1).

As shown in Table 2, the hierarchical models showed that there was a positive effect for the intervention on the outcomes linked to the Spanish achievement test. While there were positive effects for the mathematics outcomes, they were not significant.

Conclusions

The results suggest that the ETC program improves school-level aggregates of academic achievement, providing support to one of the main objectives of this nationwide educational policy implemented in Mexico. Further research must explore whether these school-level positive results are also found in student-level data.

References

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.

Gómez-Zermeño, M. G., Flores, M., Alemán, L. (2013). The full-time school program in Mexico. *Journal of Case Studies in Education, 5*(4), 1-15

Guo S., & Fraser, M. W. (2010). *Propensity Score Analysis: Statistical Methods and Applications*. Menlo Park, CA: Sage.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine, 32*(19), 3388-3414.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Ridgeway, G., McCaffrey, D., Morral, A., Griffin, B. A., & Burgette, L. (2016). *twang: Toolkit for Weighting and Analysis of Nonequivalent Groups* (Version 1.4-9.5). Retrieved from https://CRAN.R-project.org/package=twang

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Somers, M., Zhu, P., Jacob, R., & Bloom, H., (2013). *The validity and precision of the comparative interrupted time series design and the difference-in-difference design in educational evaluation* (MDRC working paper in research methodology). Retrieved from http://www.mdrc.org/publication/validity-and-precision-comparative-interrupted-time-series-design-and-difference

Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. E., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Services and Outcomes Research Methodology, 14*(4), 166-182.

Appendix

Table 1. *Descriptive statistics for the outcomes by school year*

		Comparison group				Intervention group			
		Complete sample		Matched sample		ETC not implemented		ETC implemented	
Variables	**Year**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Spanish	2006	478.91	63.15	477.22	64.30	477.05	64.93		
	2007	478.74	65.69	480.57	67.05	476.86	67.80	532.79	56.60
	2008	466.37	103.27	470.64	99.29	467.21	100.04	507.58	71.33
	2009	461.64	120.35	465.70	111.15	463.63	107.19	487.99	98.27
	2010	503.32	106.40	505.99	96.40	501.25	102.81	525.52	81.01
	2011	520.68	71.42	519.57	72.13	506.96	75.46	542.09	66.21
	2012	519.19	78.61	516.71	82.28	500.23	86.92	540.80	73.98
	2013	520.77	78.91	519.11	84.00	528.17	84.24	518.23	87.63
Mathematics	2006	482.78	66.06	481.73	66.98	479.69	67.06		
	2007	482.26	70.27	484.84	71.88	478.98	71.17	536.52	66.25
	2008	472.42	111.02	477.09	106.93	470.23	108.51	511.76	81.51
	2009	473.51	129.76	478.41	120.06	473.94	116.80	500.65	107.16
	2010	508.15	115.88	510.75	106.68	503.18	111.34	530.64	90.91
	2011	528.46	85.19	527.85	85.87	512.63	90.12	545.94	78.49
	2012	545.60	93.08	544.03	96.78	521.93	102.51	568.78	85.78
	2013	564.56	97.47	563.53	102.51	571.14	102.39	558.18	107.78
% of students	2006	26.68	25.73	27.64	26.15	27.91	26.68		
classified in	2007	27.02	26.35	26.88	26.37	28.41	27.33	13.39	14.43
the insufficient	2008	28.28	26.46	28.02	26.38	29.32	27.30	21.58	21.99
achievement	2009	27.22	26.77	28.44	27.05	30.06	28.41	24.07	23.22
level in	2010	16.66	20.24	17.81	21.36	18.36	22.49	14.98	17.57
Spanish	2011	18.61	21.54	19.33	22.11	22.30	25.66	14.91	15.65
	2012	23.24	23.77	24.39	25.13	28.87	29.07	19.02	18.77
	2013	20.36	23.21	22.04	25.05	19.85	22.60	22.90	26.40
% of students	2006	0.64	3.29	0.66	3.24	0.67	2.67		
classified in	2007	0.99	3.56	1.14	4.08	1.11	4.02	3.80	5.07
the excellent	2008	1.30	4.16	1.47	4.64	1.44	4.53	2.92	5.67
achievement	2009	1.39	4.94	1.50	5.42	1.43	4.83	2.05	5.45
level in	2010	1.66	6.07	1.73	6.10	1.69	5.91	2.65	7.18
Spanish	2011	3.90	10.22	3.91	10.19	3.19	10.09	6.28	11.40
	2012	5.65	12.11	5.90	12.95	4.71	12.63	8.57	13.59
	2013	5.35	12.91	5.79	14.00	6.90	16.08	6.26	15.08
% of students	2006	24.67	23.78	25.47	24.25	26.13	24.99		
classified in	2007	27.26	25.47	27.04	25.64	28.75	26.65	13.86	14.98
the insufficient	2008	29.56	27.07	29.57	27.08	31.50	28.25	23.37	22.48
achievement	2009	25.31	26.22	26.41	26.69	28.51	28.13	22.56	23.45
level in	2010	16.69	20.77	17.78	21.54	19.05	23.31	15.08	17.74
mathematics	2011	17.96	21.80	18.44	22.17	22.26	26.33	15.11	16.16
	2012	19.10	22.69	19.94	23.75	25.75	28.77	15.00	16.70
	2013	16.44	21.64	17.44	23.00	16.78	21.96	19.45	25.47
% of students	2006	0.57	3.70	0.56	3.47	0.62	3.93		
classified in	2007	1.54	6.26	1.77	6.96	1.59	6.22	4.71	8.22
the excellent	2008	2.27	6.78	2.44	7.02	2.27	6.51	4.08	7.90
achievement	2009	2.75	8.22	2.95	8.71	2.72	8.49	4.09	9.86
level in	2010	3.60	10.85	3.76	11.24	3.28	10.14	4.58	11.35
mathematics	2011	5.58	13.14	5.74	13.63	4.62	13.11	7.91	14.21
	2012	9.19	16.53	9.37	17.26	7.40	16.75	12.31	17.20
	2013	11.64	19.34	12.23	20.67	13.08	20.94	12.47	20.99

Note: SD refers to standard deviation.

Table 2. *Estimated coefficients for the ETC effect on the outcomes*

	Complete sample (N = 62,246)			**Matched sample (N = 21,400)**			**Matched sample with multiple imputation**		
Effects	**Coef**	**S.E.**	***p-value***	**Coef**	**S.E.**	***p-value***	**Coef**	**S.E.**	***p-value***
Test score in Spanish									
Intercepto, γ_{00}	469.07	2.71	<0.001	562.08	7.49	<0.001	559.51	7.88	<0.001
T_i, γ_{10}	29.79	7.55	**<0.001**	5.56	9.35	0.55	7.56	9.93	0.45
ETC_{ti}, γ_{11}	-2.34	7.11	0.74	21.04	8.45	**0.01**	25.64	9.08	**0.01**
% of students classified in the insufficient achievement level in Spanish									
Intercepto, γ_{00}	30.33	0.80	<0.001	3.64	2.01	0.07	3.19	2.02	0.11
T_i, γ_{10}	-2.43	3.82	0.52	10.56	6.53	0.11	10.78	6.56	0.10
ETC_{ti}, γ_{11}	-4.86	3.74	0.19	-16.40	6.43	**0.01**	-16.40	6.46	**0.01**
% of students classified in the excellent achievement level in Spanish									
Intercepto, γ_{00}	2.01	0.25	<0.001	7.04	0.56	<0.001	7.05	0.55	<0.001
T_i, γ_{10}	-0.49	0.40	0.22	-0.23	0.40	0.57	-0.29	0.42	0.48
ETC_{ti}, γ_{11}	1.01	0.34	**0.00**	0.83	0.27	**0.00**	0.85	0.30	**0.01**
Test score in mathematics									
Intercepto, γ_{00}	479.85	2.97	<0.001	535.47	8.66	<0.001	536.58	8.57	<0.001
T_i, γ_{10}	23.91	9.75	**0.01**	-3.85	17.36	0.83	-2.83	16.04	0.86
ETC_{ti}, γ_{11}	-8.56	9.36	0.36	30.75	16.79	0.07	28.47	15.44	0.07
% of students classified in the insufficient achievement level in mathematics									
Intercepto, γ_{00}	22.90	0.83	<0.001	4.42	2.14	0.04	4.17	2.14	0.05
T_i, γ_{10}	-0.72	3.93	0.85	12.43	9.50	0.19	12.41	9.40	0.19
ETC_{ti}, γ_{11}	-0.23	3.85	0.95	-14.01	9.43	0.14	-13.83	9.32	0.14
% of students classified in the excellent achievement level in mathematics									
Intercepto, γ_{00}	1.85	0.29	<0.001	4.84	0.77	<0.001	4.95	0.77	<0.001
T_i, γ_{10}	0.91	1.11	0.42	0.28	0.50	0.57	0.31	0.53	0.57
ETC_{ti}, γ_{11}	-0.37	1.09	0.74	0.37	0.34	0.27	0.23	0.38	0.55

Note: *Coef* refers to the estimated coefficient in the hiearchical model, *S.E.* to the standard error of the coefficient.