

Impact of English instruction on labor market outcomes

The case of Mexico

Oscar Galvez-Soriano

University of Houston
Department of Economics

May, 2022

Motivation

The value of English language skills in developing non-English speaking countries

- Globalization: trade and culture (internet, news, social media, etc.)
- Migration and labor market outcomes

I will study the expansion of English instruction in Mexico

Related Literature

- Returns to English language skills
 - In English-speaking countries: Chiswick and Miller (1995); Bleakley and Chin (2004)
 - In non English-speaking countries: Lang and Siniver (2009); Azam, Chin and Prakash (2013); Chakraborty and Bakshi (2016)
- English instruction in schools
 - Policy change in the medium of instruction: Angrist, Chin and Godoy (2008); Eriksson (2014)

I study English instruction as a subject in Mexican public primary schools

Main contributions of this paper

Research Question

- What is the effect of exposure to English instruction on labor market outcomes in a non-English speaking country?

Empirical strategy

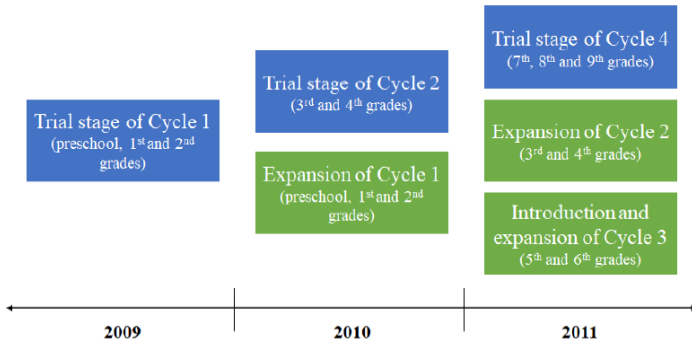
- Variation in exposure driven by a policy change in Mexico
- Rich data connecting individuals in primary school to their labor market outcomes
- Use school by cohort variation in exposure to English instruction in Mexican primary schools

Policy background

National English Program in Basic Education (NEPBE)
launched in 2009 in Mexico

- Introduced English instruction in public primary schools
- Funded by the central government
- Implemented gradually

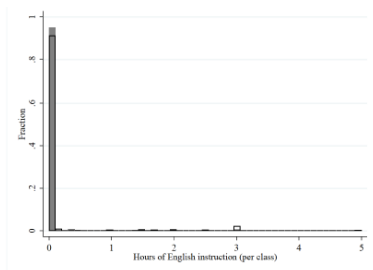
Policy background: English program stages



Note: NEPBE was launched in 2009 as a trial stage with the called Cycle 1. In 2010 the program continued the trial stage with the Cycle 2 and expanded Cycle 1. Finally, in 2011 the program introduced for the first time and expanded Cycle 3, benefiting fifth and sixth graders.

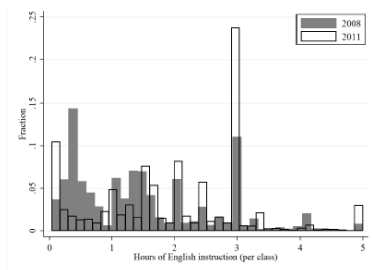
» data

Policy background: Distribution of hours of English instruction (2008 vs 2011)



(a) Hours of English instruction

Note: Density of the indicated variables are plotted. Histograms at the right do not show zeros, which capture most of the distribution. Hours of English instruction are calculated dividing total hours in a school by total number of classes. Similarly, number of English teachers are calculated dividing total number of English teachers by total number of classes in a school.



(b) Hours of English instruction (w/o zeros)

Does exposure to English instruction increase English abilities?

In 2008, the Mexican Think Tank CIDAC conducted a survey to measure English abilities among individuals (15-55) living in urban areas

- About 35% of the urban population understand English
- Among these individuals, 75% acquired their English abilities in school

Mexican Subjective Well-being Survey (BIARE) conducted in 2014 by the Mexican Institute of Statistics (INEGI):

- High no-response rate to English ability question; 54% who responded said they speak English

Empirical strategy

- Challenging to estimate the effect of exposure to English instruction on labor market outcomes
- Key concern: schools that offered English instruction are systematically different from those that did not
 - Likely to have positive selection bias, e.g., schools offering English instruction located in richer neighborhoods
- I address this by using a school FE approach
 - Data of the universe of primary school students, able to connect to their labor market outcomes
 - Intuition: compare students from the same school, some with more English instruction and some with less

Measure of exposure to English instruction

Cohort	Primary school					
	1st	2nd	3rd	4th	5th	6th
1997	2003	2004	2005	2006	2007	2008
1998	2004	2005	2006	2007	2008	2009
1999	2005	2006	2007	2008	2009	2010
2000	2006	2007	2008	2009	2010	2011
2001	2007	2008	2009	2010	2011	2012
2002	2008	2009	2010	2011	2012	2013

- Using the Mexican school census, I calculate weekly hours of English instruction (per class), for each school-year
- For each school-cohort, I average the hours of English instruction from 1st to 6th grade
 - I assume students enter school at age 6 and had normal progression until grade 6th

Impact on labor market outcomes

I estimate the following equation to get the effect of exposure to English instruction on labor market outcomes:

$$y_{isct} = \alpha + \beta \cdot ExpEng_{sc} + \mathbf{X}_{isct}\boldsymbol{\gamma} + \zeta_c + \nu_s + \tau_t + \varepsilon_{isct}$$

where y_{isct} is the labor market outcome of individual i , who attended school s , belongs to cohort c and observed at time t

Mechanisms

Potential mechanisms of the effect of exposure to English instruction on labor market outcomes:

- English language skills
- Other skills
 - Language (Spanish)
 - Mathematics
- Potential heterogeneity by abilities

Data

I use three main sources of data:

- Social Security data (2018-2021)
 - Worker level
 - Formal sector
 - Individuals between 16-22 years old
- ENLACE (2006-2013): Nationwide test (Math and Language)
 - Student level
 - From 3rd to 6th grade
- Mexican school census (2003-2013)
 - School level
 - School characteristics: weekly hours of English instruction

Data: Labor market outcomes

I investigate the effect of exposure to English instruction on four main labor market outcomes:

- ① Probability of working in formal sector
 - Dummy for being in social security data among the universe of students
- ② Wages (average monthly wage)
- ③ Geographical mobility (distance from home county to working county)
- ④ Economic industries (NAICS) [» codes](#)

Estimation results: exposure to English instruction and labor market outcomes

Table 2: Exposure to English instruction and labor market outcomes (Social Security data)

	(1) Formal Sector	(2) ln(wage)	(3) ln(distance)	(4) Move State
<i>Panel A: Full sample</i>				
Hrs English	-0.012*** (0.001)	-0.014*** (0.002)	-0.033*** (0.008)	-0.004*** (0.001)
Observations	15,137,341	3,588,164	3,588,164	3,588,164
Adjusted R^2	0.108	0.239	0.478	0.556

- Concern about selection into social security data
- Possibly because individuals are still enrolled in school
- Use counties where it is less likely that they are enrolled

Proposed solution: Construction of low-enrollment sample

- ① Using the 2020 Mexican Population Census, I construct a county-enrollment rate variable
 - Enrollment rates in first year of college (2002 cohort)
- ② I keep the data with 38 percent (or less) of individuals enrolled in school

►► How?

Labor market outcomes with low-enrollment sample

Table 2: Exposure to English instruction and labor market outcomes (Social Security data)

	(1)	(2)	(3)	(4)
	Formal Sector	ln(wage)	ln(distance)	Move State
<i>Panel B: Low enrollment sample</i>				
Hrs English	-0.012 (0.008)	-0.000 (0.011)	-0.075* (0.045)	0.013 (0.009)
Observations	1,388,210	232,187	232,187	232,187
Adjusted R^2	0.125	0.293	0.679	0.727
<i>Panel C: Low enrollment sample (Men)</i>				
Hrs English (β^M)	-0.018 (0.011)	-0.002 (0.017)	-0.142** (0.059)	0.003 (0.013)
Observations	669,325	148,595	148,595	148,595
Adjusted R^2	0.151	0.296	0.681	0.729
<i>Panel D: Low enrollment sample (Women)</i>				
Hrs English (β^W)	-0.008 (0.010)	-0.009 (0.015)	0.037 (0.068)	0.027** (0.013)
Observations	718,885	83,592	83,592	83,592
Adjusted R^2	0.110	0.347	0.703	0.756
$\beta^M = \beta^W$ [p-value]	[0.015]	[0.561]	[0.063]	[0.109]
State of work FE	NO	YES	YES	YES

Labor market outcomes with low-enrollment sample

Table 2: Exposure to English instruction and labor market outcomes (Social Security data)

	(1)	(2)	(3)	(4)
	Formal Sector	ln(wage)	ln(distance)	Move State
<i>Panel B: Low enrollment sample</i>				
Hrs English	-0.012 (0.008)	-0.000 (0.011)	-0.075* (0.045)	0.013 (0.009)
Observations	1,388,210	232,187	232,187	232,187
Adjusted R^2	0.125	0.293	0.679	0.727
<i>Panel C: Low enrollment sample (Men)</i>				
Hrs English (β^M)	-0.018 (0.011)	-0.002 (0.017)	-0.142** (0.059)	0.003 (0.013)
Observations	669,325	148,595	148,595	148,595
Adjusted R^2	0.151	0.296	0.681	0.729
<i>Panel D: Low enrollment sample (Women)</i>				
Hrs English (β^W)	-0.008 (0.010)	-0.009 (0.015)	0.037 (0.068)	0.027** (0.013)
Observations	718,885	83,592	83,592	83,592
Adjusted R^2	0.110	0.347	0.703	0.756
$\beta^M = \beta^W$ [p-value]	[0.015]	[0.561]	[0.063]	[0.109]
State of work FE	NO	YES	YES	YES

Labor market outcomes with low-enrollment sample

Table 2: Exposure to English instruction and labor market outcomes (Social Security data)

	(1)	(2)	(3)	(4)
	Formal Sector	ln(wage)	ln(distance)	Move State
<i>Panel B: Low enrollment sample</i>				
Hrs English	-0.012 (0.008)	-0.000 (0.011)	-0.075* (0.045)	0.013 (0.009)
Observations	1,388,210	232,187	232,187	232,187
Adjusted R^2	0.125	0.293	0.679	0.727
<i>Panel C: Low enrollment sample (Men)</i>				
Hrs English (β^M)	-0.018 (0.011)	-0.002 (0.017)	-0.142** (0.059)	0.003 (0.013)
Observations	669,325	148,595	148,595	148,595
Adjusted R^2	0.151	0.296	0.681	0.729
<i>Panel D: Low enrollment sample (Women)</i>				
Hrs English (β^W)	-0.008 (0.010)	-0.009 (0.015)	0.037 (0.068)	0.027** (0.013)
Observations	718,885	83,592	83,592	83,592
Adjusted R^2	0.110	0.347	0.703	0.756
$\beta^M = \beta^W$ [p-value]	[0.015]	[0.561]	[0.063]	[0.109]
State of work FE	NO	YES	YES	YES

Labor market outcomes by abilities

Table A.10: Exposure to English instruction and labor market outcomes by abilities
(Social Security data)

	(1)	(2)	(3)	(4)
	Formal Sector	ln(wage)	ln(distance)	Move State
<i>Panel A: Low enrollment sample</i>				
Hrs English	-0.008 (0.009)	-0.009 (0.012)	-0.102** (0.050)	0.018* (0.011)
Eng×Q2	-0.002 (0.006)	-0.002 (0.008)	0.004 (0.049)	-0.008 (0.009)
Eng×Q3	-0.004 (0.006)	0.032*** (0.009)	0.013 (0.038)	-0.017 (0.011)
Eng×Q4	-0.015** (0.007)	0.009 (0.012)	0.109*** (0.041)	0.002 (0.012)
Observations	1,388,210	232,187	232,187	232,187
Adjusted R^2	0.126	0.293	0.679	0.727

English instruction and economic industries

Table 3: Exposure to English instruction and economic industries (Social Security data)

	(1)	(2)	(3)	(4)
	Agri- culture	Con- struction	Manu- facture	Serv- ices
<i>Panel B: Low enrollment sample</i>				
Hrs English	-0.010 (0.006)	-0.024** (0.010)	0.036** (0.017)	-0.001 (0.017)
Observations	232,309	232,309	232,309	232,309
Adjusted R^2	0.407	0.391	0.345	0.294
<i>Panel C: Low enrollment sample (Men)</i>				
Hrs English (β^M)	-0.002 (0.009)	-0.027* (0.015)	0.033 (0.021)	-0.004 (0.020)
Observations	148,654	148,654	148,654	148,654
Adjusted R^2	0.427	0.426	0.353	0.274
<i>Panel D: Low enrollment sample (Women)</i>				
Hrs English (β^W)	-0.024*** (0.008)	-0.004 (0.006)	0.045** (0.023)	-0.017 (0.025)
Observations	83,655	83,655	83,655	83,655
Adjusted R^2	0.451	0.144	0.388	0.386
$\beta^M = \beta^W$ [p-value]	[0.888]	[0.000]	[0.002]	[0.779]
Shares	0.04	0.08	0.35	0.53

Additional outcome measures based on industry codes

- Currently working on constructing measures of industry based on English intensity
- Take advantage of BIARE data set
 - Construct a share of workers who report speaking English (by industry 4-digit code, NAICS)
- Jobs that have high share workers may not necessarily be jobs that use English
 - I also will use extent to which job ads in Mexico seek bilingual speakers (Delgado-Helleseter, 2020)

Summary of results so far

- Looking at individuals 16-22 year old, I find that more exposure to English instruction makes them less likely to work in formal labor
- I use a low-enrollment sample that deals with selection into formal sector
 - Find no average effect on wages
 - However, for high ability individuals there is a positive effect
 - Movements from agriculture and construction to manufactures

Exploring mechanisms

- Mechanism 1: English abilities
 - I use the 2014 Subjective Well-being Survey (BIARE)
 - Representative at national and state level
 - Asks if the respondent knows how to speak English
- Mechanism 2: Cognitive abilities
 - Test score data
 - School FE approach

Mechanism 2: Cognitive abilities

Effect of exposure to English instruction on student achievement:

$$test_score_{isc} = \theta + \phi \cdot ExpEng_{sc} + \mathbf{X}_{isc}\boldsymbol{\gamma} + \zeta_c + \nu_s + \varepsilon_{isc}$$

where $test_score_{isc}$ is the 6th grade test score of student i , who attended school s and belongs to cohort c

» Data

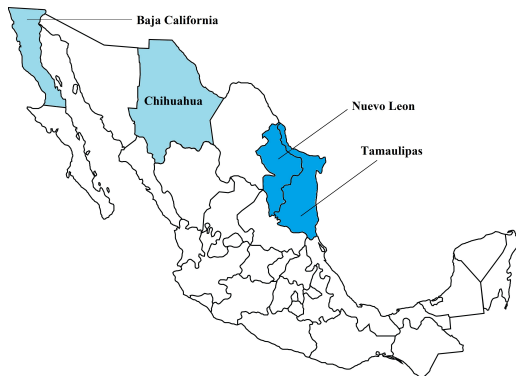
Estimation results: exposure to English instruction and test scores

Table 7: Exposure to English instruction and student achievement

	(1)	(2)	(3)	(4)
	Language 6th	Language 6th	Math 6th	Math 6th
<i>Panel A: Full sample in Social Security data</i>				
Hrs English	0.0316*** (0.0033)	-0.0015 (0.0075)	0.0133*** (0.0037)	-0.0225*** (0.0086)
Observations	3,588,164	3,588,164	3,588,164	3,588,164
Adjusted R^2	0.396	0.447	0.402	0.462
<i>Panel B: Low enrollment sample</i>				
Hrs English	0.0496 (0.0469)	0.0548 (0.0867)	0.0124 (0.0343)	0.0128 (0.0652)
Observations	232,187	232,187	232,187	232,187
Adjusted R^2	0.342	0.437	0.373	0.472
State FE	YES	NO	YES	NO
School FE	NO	YES	NO	YES

Mechanism 1: English abilities

The Mexican states of Nuevo Leon and Tamaulipas introduced English programs in 1998 and 2002, respectively, to offer English instruction in elementary school



- Cohorts 1987-1996 had exposure to Eng instruction in elementary schools of Nuevo Leon
- Cohorts 1991-1996 had exposure to Eng instruction in elementary schools of Tamaulipas

Mechanism 1: English abilities

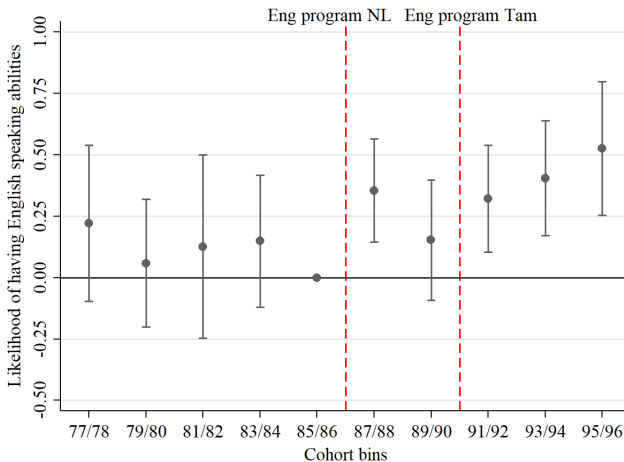
I estimate the intention to treat effect of offering English instruction in elementary school on English abilities (eng_{isc}) using a difference in differences approach

- $HadPolicy_{sc}$: takes the value of one if the individual i lives in Nuevo Leon and he/she was born since 1987, or
- if individual i lives in Tamaulipas and he/she was born since 1991; zero otherwise

$$eng_{isc} = \alpha + \beta \cdot HadPolicy_{sc} + \delta_s + \gamma_c + \mathbf{X}_{isc}\boldsymbol{\lambda} + \varepsilon_{isc}$$

$$eng_{isc} = \alpha + \sum_c \beta_c \cdot I_{(treatment_{sc}=c)} + \delta_s + \gamma_c + \mathbf{X}_{isc}\boldsymbol{\lambda} + \varepsilon_{isc}$$

English instruction increases English abilities



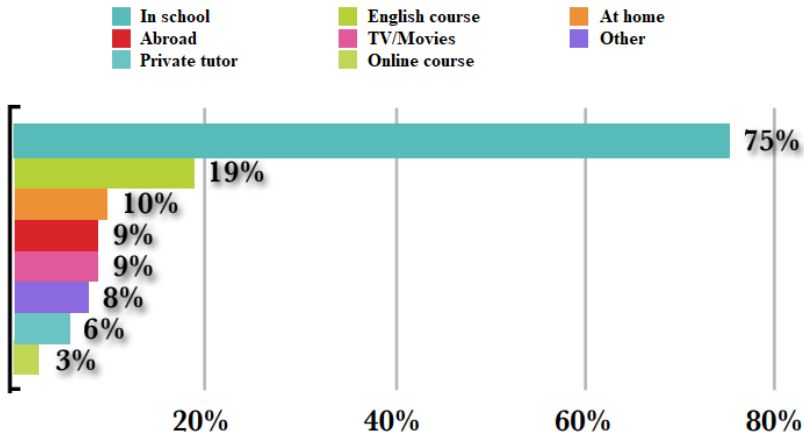
Next steps

- Explore alternative sources of heterogeneity
- School input data
- Enrollment in public relative to private primary schools
- Work with the Mexican Population Census data
 - Education
 - International Migration
 - Occupations

Thank you!

Oscar Galvez-Soriano
University of Houston
(ogalvezs@cougarnet.uh.edu)

Where did you learn English?



Note: This question was answered only by individuals who reported having English abilities. The answers are independent, i.e. do not sum 100 percent.

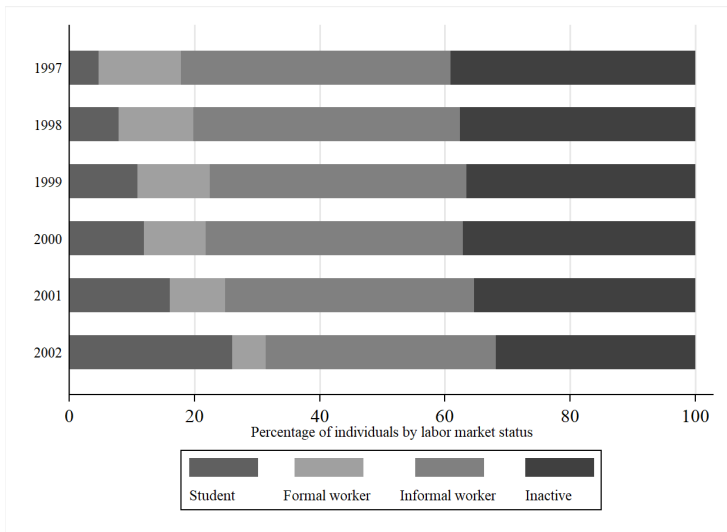
Source: CIDAC (2008). Encuesta CIDAC sobre Capital Humano en México. México.

Descriptive statistics (matched data sets)

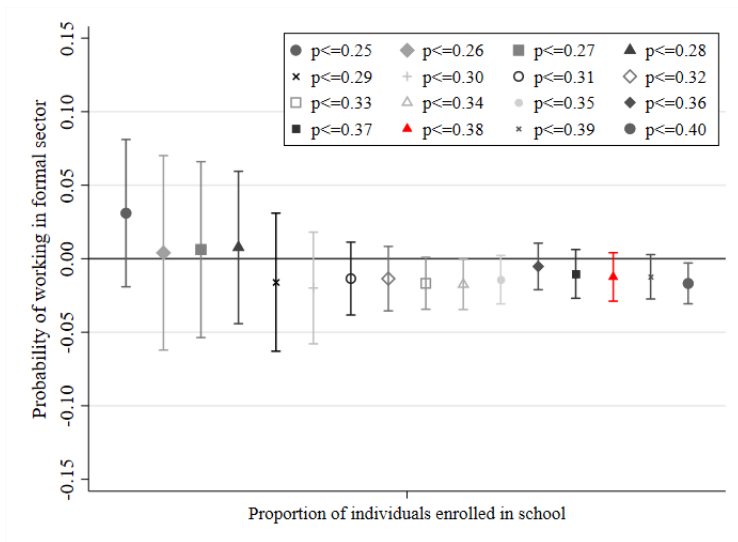
Table 1: Descriptive statistics

Variable	Mean	SD	Min	Max
<i>Individual characteristics</i>				
Female	0.39	0.49	0	1
Age	20.42	1.27	16	22
Language test score	-0.06	0.97	-2.84	3.49
Math test score	-0.05	0.97	-2.69	3.40
<i>School characteristics</i>				
Hours of English instruction	0.26	0.70	0	29.18
English teachers	0.02	0.06	0	1.33
Number of students (6th grade)	28.82	9.51	1	119
Number of teachers with college	0.86	0.21	0	3
Number of teachers with masters	0.06	0.08	0	1
Rural	0.27	0.44	0	1
<i>Labor characteristics</i>				
Wage (monthly pesos)	6,038	2,862	2,510	65,160
Permanent	0.81	0.39	0	1
Number of jobs (in a year)	1.43	0.79	1	17
Number of permanent jobs	1.15	0.80	0	14
Distance home-work (km)	106.53	261.15	0	2,029
Observations	2,999,443			

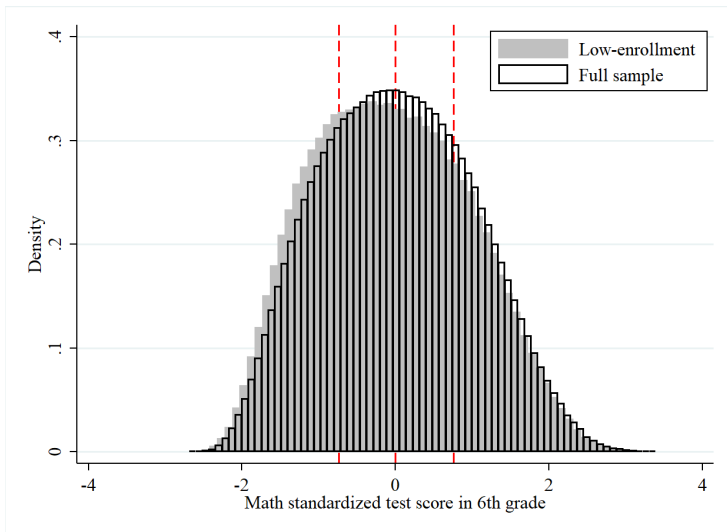
Statuses with low-enrollment sample



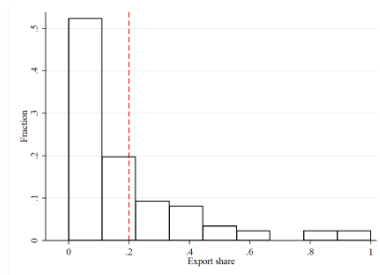
How I chose the 38 percent enrollment rate?



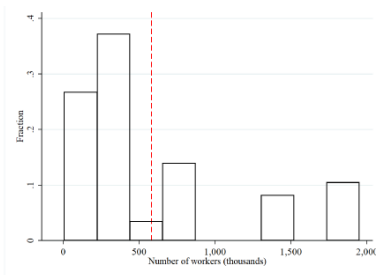
Abilities distribution (full vs. low-enrollment sample)



Manufacturing industries with high-export share



(a) Exports share



(b) Number of workers

Note: Distributions of exports share and number of workers are plotted. Each observation in the distribution corresponds to an economic industry in the sub-sector of manufactures. I use the North American Industry Classification System (NAICS) at four-digit code to classify each industry, which means that I work with a total of 86 different economic industries.

» Back

Manufacturing industries with high-export share

- Seafood Product Preparation and Packaging
- Apparel Accessories
- Forging and Stamping
- Hardware Manufacturing
- Spring and Wire Product
- Other Fabricated Metal Product
- Motor Vehicle
- Motor Vehicle Body and Trailer
- Motor Vehicle Parts
- Ship and Boat Building

[» Back](#)

Services that require English abilities

Eng abilities	Services from NAICS classification
0.0189	Truck Transportation
0.0191	Transportation Equipment Manufacturing
0.0194	Sporting Goods, Hobby, Musical Instrument, Book, and Miscellaneous Retailers
0.0195	Transit and Ground Passenger Transportation
0.0198	Religious, Grantmaking, Civic, Professional, and Similar Organizations
0.0202	Executive, Legislative, and Other General Government Support
0.0226	Drugs and Druggists' Sundries Merchant Wholesalers
0.0232	Administrative and Support Services
0.0249	Newspaper, Periodical, Book, and Directory Publishers
0.0254	Printing and Related Support Activities
0.0274	Merchant Wholesalers, Durable Goods
0.0322	Credit Intermediation and Related Activities
0.0352	Museums, Historical Sites, and Similar Institutions
0.0361	Air Transportation
0.0383	Performing Arts, Spectator Sports, and Related Industries
0.0393	Amusement, Gambling, and Recreation Industries
0.0653	Support Activities for Transportation
0.0675	Real Estate
0.0763	Professional, Scientific, and Technical Services
0.0791	Warehousing and Storage
0.0893	Accommodation
0.1269	Scenic and Sightseeing Transportation

NAICS codes in my classification

Industries	NAICS code	Industry Title
Agriculture	11	Agriculture, Forestry, Fishing and Hunting
	21	Mining
Construction	22	Utilities
	23	Construction
Manufacturing	31-33	Manufacturing
	42	Wholesale Trade
	44-45	Retail Trade
	48-49	Transportation and Warehousing
	51	Information
	52	Finance and Insurance
	53	Real Estate Rental and Leasing
Services	54	Professional, Scientific, and Technical Services
	55	Management of Companies and Enterprises
	56	Administrative and Support and Waste Management
	61	Educational Services
	62	Health Care and Social Assistance
	71	Arts, Entertainment, and Recreation
	72	Accommodation and Food Services
	81	Other Services (except Public Administration)
	92	Public Administration

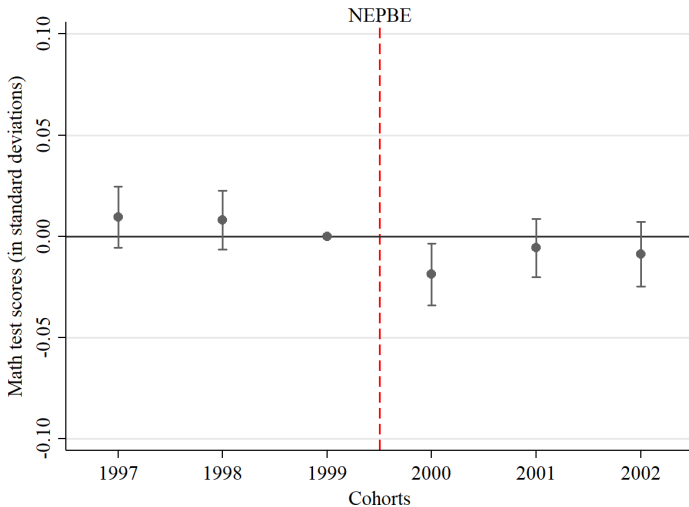
Pre-trends (empirical strategy)

To rule out the existence of different pre-trends among cohorts, I use an event study type regression (where the cohort 1999 is the reference group), as follows:

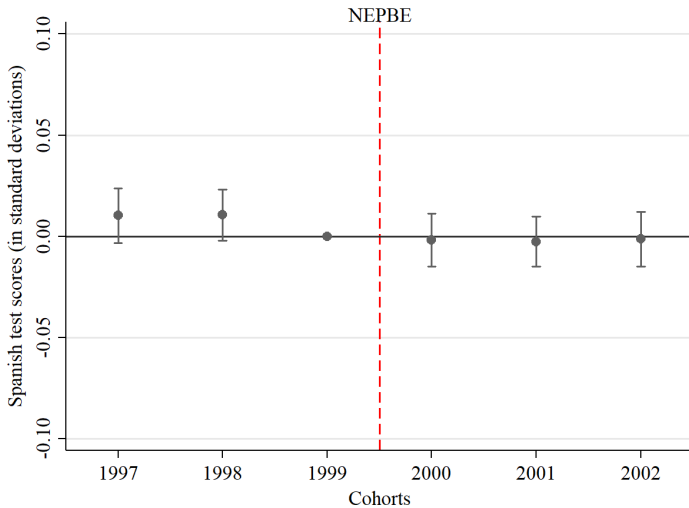
$$\begin{aligned} test_score_{isc} = & \alpha + \sum_c \beta_c \cdot ExpEng_{sc} \times I_{(cohort_c=c)} + \mathbf{X}_{isc} \gamma \\ & + \zeta_c + \nu_s + \varepsilon_{isc} \end{aligned}$$

where ζ_c controls for cohort FE, ν_s accounts for school FE, and $I_{(cohort_c=c)}$ is an indicator function with $C = \{1997, 1998, 2000, 2001, 2002\}$

Pre-trends Math test scores



Pre-trends Spanish test scores



Does exposure to English instruction increase English abilities? The case of Nuevo Leon

The Mexican state of Nuevo Leon introduced an English program in 1998 to offer English instruction only in the last year of elementary school (sixth grade)



- Not all elementary schools were beneficiaries
- Cohorts 1987 (and younger) had exposure to one year of English instruction in elementary school

Nuevo Leon Eng Program / Empirical strategy

I estimate the intention to treat effect of offering one year of English instruction in elementary school on English abilities (eng_{isc}) using a difference in differences approach

- $treatment_s$: takes the value of one if the individual i lives in Nuevo Leon, zero if he/she lives in Chihuahua
- $after_c$: takes the value of one for cohorts (1987-1996) affected by the policy, and zero for non-affected cohorts (1977-1986)

$$eng_{isc} = \alpha + \beta (treatment_s \times after_c) + \delta \cdot treatment_s + \gamma \cdot after_c + \mathbf{X}_{isc}\boldsymbol{\lambda} + \varepsilon_{isc}$$

Nuevo Leon English Program / Data

- Subjective well-being survey (BIARE, 2014)
- Representative at national and state level
- Respondents are adults 18 years of age and older
- This survey asks if the respondent knows how to speak in English

English instruction increases English abilities / PTA

To assess the validity of the parallel trend assumption, I use the cohorts 1977-1981 as the omitted group in an event study type regression, as follows:

$$eng_{isc} = \alpha + \sum_C \beta_C I_{(cohort_c=C)} + \delta treatment_s + \gamma_c \\ + \mathbf{X}_{isc} \boldsymbol{\lambda} + \varepsilon_{isc}$$

where $I_{(cohort_c=C)}$ is an indicator function for cohort bins with $C = \{1982 - 1986, 1987 - 1991, 1992 - 1996\}$

English instruction increases English abilities

Table : English instruction and formation of English abilities (Nuevo Leon)

	(1)	(2)	(3)	(4)
	Eng	Eng	Eng	Eng
After×Treat	0.342** (0.129)	0.383*** (0.099)	0.520*** (0.108)	0.507*** (0.145)
Observations	110	110	110	110
Adjusted R^2	0.033	0.588	0.617	0.694
State FE	YES	YES	YES	NO
Controls	NO	YES	YES	YES
Cohort FE	NO	NO	YES	YES
County FE	NO	NO	NO	YES

Data: Student achievement

- 1 I look at test scores as one of the mechanisms
- 2 I standardize test scores, ts_{isct} , of each student i in school s at time t using the following formula:

$$test_score_{ist} = \frac{ts_{isct} - \mu_t}{\sigma_t}$$

where $test_score_{ist}$ is the standardized test score, while μ_t and σ_t are the mean and standard deviation of test scores, respectively, pooling all Mexican students by grade and by each observed year

» Back