Impact of English instruction on labor market outcomes: The case of Mexico

Oscar Gálvez-Soriano*
University of Houston

March 2022

Abstract

In this paper I measure the effect of being exposed to a foreign language in school on labor market outcomes and student achievement. I exploit a policy change in Mexico that introduced English instruction in elementary schools through the English Program in Basic Education (NEPBE) in 2009. I construct a novel database, which contains nation-wide information of elementary school students linked to school panel data on characteristics like hours of English instruction as well as to their labor market records in adulthood. Using a school fixed effects approach, I find that exposure to English instruction reduces the likelihood that an individual participates in formal sector employment. It is likely that this result is due to exposure to English instruction affecting enrollment in high school and college, as my analysis focuses on young adults aged 16-22. Focusing on a sub-sample that is unlikely to be enrolled by age 16, I find that the exposure to English instruction had no effect on wages. I also evaluate the effect of exposure to English instruction on students' achievement to determine if part of the effect on wages is due to a reallocation of resources towards English instruction in primary schools, which can potentially affect the formation of human capital. I find no effects on Language and Math test scores, which suggests that the estimated effect of exposure to English language on wages is not reflecting changes to general cognitive skills.

JEL Classification: I21, I28, J24, O15.

Keywords: Early Childhood Education, Education Reform, Skills, Education Policy.

*Department of Economics, 3623 Cullen Boulevard, office 250 McElhinney Hall. Houston, TX 77204. e-mail: ogalvezs@cougarnet.uh.edu

Introduction

In the context of a globalizing world in which English is the lingua franca, there has been a growing interest among developing countries[1] in offering English instruction at school under the assumption that English instruction will increase English-language skills and ultimately improve labor market outcomes. While there is a large literature on the effect of English skills on earnings in English-speaking countries, there is very little research on the returns to English skills in non-English-speaking countries.

In this research, I offer some of the first empirical evidence about the effect of the exposure to English instruction in primary school on labor market outcomes in a non-English-speaking context. The set of labor market outcomes I examine include wages, occupations, and geographical mobility. To do that, I exploit the implementation of a program in Mexico (in 2009) that significantly expanded English instruction in Mexican primary schools called the National English Program in Basic Education (NEPBE).

I construct a novel database, which contains all public-school students in Mexico linked to information about the elementary school they attended (including hours of English instruction) as well as to their labor market records in adulthood. I document that after 2009 many English teachers were hired, and weekly hours of English instruction increased as a result of the program.

Unlike the existing literature, I do not evaluate English abilities because this is an unobservable variable in Mexico, instead I observe exposure to English instruction. Among this related literature, there is a consensus that English language skills are positively associated with immigrants' earnings in English-speaking countries (McManus, Gould and Welch, 1983; Grenier, 1984; Rivera-Batiz, 1990; Chiswick, 1991; Chiswick and Miller, 1995; Bleakley and Chin, 2004; Borjas and Friedberg, 2009; Bleakley and Chin, 2010; Chiswick and Miller, 2015). Similar results have been found in the context of immigrants to European countries (Dustmann, 1994; Dustmann and Soest, 2001; Hayfron, 2001; Williams, 2011). Likewise, it has been found that the English premium among immigrants could vary depending on the workers' age and education, i.e. younger and more educated immigrants have greater returns to English at work (Lang and Siniver, 2009; Azam, Chin and Prakash, 2013).

Although I do not observe English ability among the individuals in the administrative data, it is possible to argue that the effect of the English program on labor market outcomes is due to an improvement on English abilities and not through other skills that affect human capital, such as language or Math. I evaluate this latter argument by analyzing the effect of the NEPBE on school achievement. My results suggest that this English program had no effect neither on language nor on Math test scores of Mexican students.

To my knowledge, this paper is the first research on the effect of being exposed to English

[1] India, South Africa and Latin American countries.

instruction (as a subject) at early stages of life on labor market outcomes in a Latin American economy.[2] A related research was conducted by Delgado Helleseter (2020) to estimate the effect of English abilities in the Mexican market of jobs posted online. This author finds that the wage premium for Mexican English speakers is approximately 28 percent higher than non-English speakers. However, his sample is composed only by online advertisements from a single online job board, which implies concerns about the selection bias and the external validity of his results. Two other related papers are Azam, Chin and Prakash (2013) and Chakraborty and Bakshi (2016) in the context of India, where the authors find that English-language skills are strongly positively associated with earnings.

This paper is also different from previous research where authors exploit policy changes in the language of instruction (Angrist and Lavy, 1997; Angrist, Chin and Godoy, 2008; Eriksson, 2014; Taylor and von Fintel, 2016). Indeed, I analyze a non-English speaking country, where the language of instruction has not changed but the exposure to English increases at early stages of life through the inclusion of English language as a subject in primary schools. Besides, the setting I study in this paper is Mexico where the value of English language is even more pronounced than in other countries because of its close relationship with the United States (US) in terms of trade and migration. Furthermore, this paper is also different from previous research in non-English speaking countries that were England colonies, where the upper-social class and government workers usually have English abilities, such as in India and South Africa. Indeed, English is a much less common language in most of the non-English speaking countries.

Thus, in this paper I would like to determine if the inclusion of English in the regular curricula of primary schools could affect students achievement and, later in life, their labor market outcomes (in terms of wages, informality and economic industries). Hence, I measure the effect of exposure to English instruction at early stages of life. However, the way to answer this question is not trivial because of the selection problem. Indeed, it could be the case that students who were exposed to English instruction studied in better schools or live in better neighborhoods. I get around this selection problem by exploiting the richness of my data set, controlling for school fixed effects.

Indeed, using a school fixed effects approach, which controls for time-invariant characteristics of schools and local neighborhoods that may be correlated with the exposure to English instruction for individuals and their labor market opportunities, I find that exposure to English instruction reduces the likelihood that an individual participates in formal sector employment. It is likely that this result is due to exposure to English instruction affecting enrolling in high school and college, as my analysis focuses on young adults aged 16-22 (the recency of the NEPBE means the affected cohorts are still young).

Focusing on a subsample that is unlikely to be enrolled by age 16, I find that the exposure

[2] I am aware of a recent research in Mexico where Charles-Leija and Torres (2022) offer an estimate of the returns to English skills. However, their model has serious problems of identification because they do not solve the selection problem due to the differences between individuals who report being English speakers and those who do not.

to English instruction increases wages for men but not for women. This gender heterogeneity can be explained by a reallocation of workers in certain economic industries. I find that, after the exposure to English instruction, men increase their probability of working in export industries jobs, while women are more likely to work in hospitality and telecommunications.

I also evaluate the effect of exposure to English instruction on students' achievement to determine if part of the effect on wages is due to a reallocation of resources towards English instruction in primary schools, which can potentially affect the formation of human capital. I find no effects on Language and Math test scores, which suggests that the estimated effect of exposure to English language on wages is not reflecting changes to general cognitive skills. These findings are consistent with exposure to English instruction raising wages through increasing English language skills.

The remaining of this paper proceeds as follows. In the first section I explain the background of the policy. In section 2, I explain the empirical strategy I propose. In section 3, I describe the novel data base I construct. In section 4, I show the results of the effect of the English program on labor market outcomes and on student achievement. Finally, section 4.3 summarizes with a discussion of my findings and a brief conclusion.

1 Background

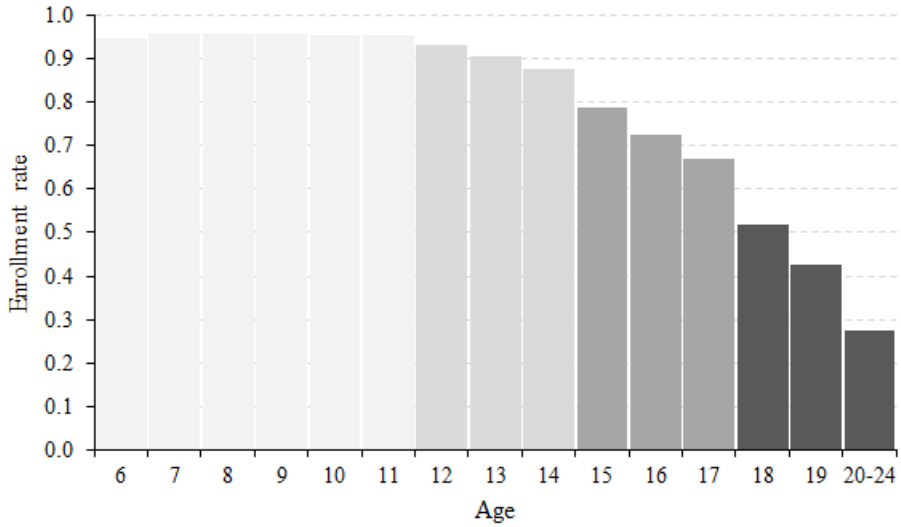
1.1 The Mexican education system and labor participation

In Mexico, the basic education is divided in two levels: elementary and middle school (or primary and lower secondary, respectively). The elementary schools comprise from first to sixth grade, while middle schools comprise from seventh to ninth grade. Both educational levels are part of the basic compulsory education system as it also is the high school (since 2013), which comprises from tenth to twelfth grade.

Regularly, students enrolled in elementary school range of ages from six to eleven years old (at the beginning of the second term of each academic year). Students enrolled in middle school are between 12 and 14 years of age, while students enrolled in high school are between 15 and 17 years old (at the beginning of the second term of each academic year as well). College students are usually between 18 and 21 years of age.

Enrollment rates in primary school are close to one hundred percent, but in college the enrollment rate falls dramatically. In Figure 1 I plot enrollment rates by age using data from the 2020 Mexican population census. To interpret these data as enrollment rates per grade I assume that students who reported attending school are enrolled in the grade that corresponds to their age. Thus, the different bars colors represent different levels of education. There are several factors that could explain the decreasing enrollment rates in higher academic

Figure 1: Enrollment rates by age (based on 2020 Mexican census data).



Note: Enrollment rates by age are plotted. I assume that students who reported attending school are enrolled in the grade that corresponds to their age. Hence, the first six light gray bars represent primary school enrollment, the next three darker bars represent middle school enrollment, high school and, finally, the dark gray bars represent college and graduate degrees enrollment.

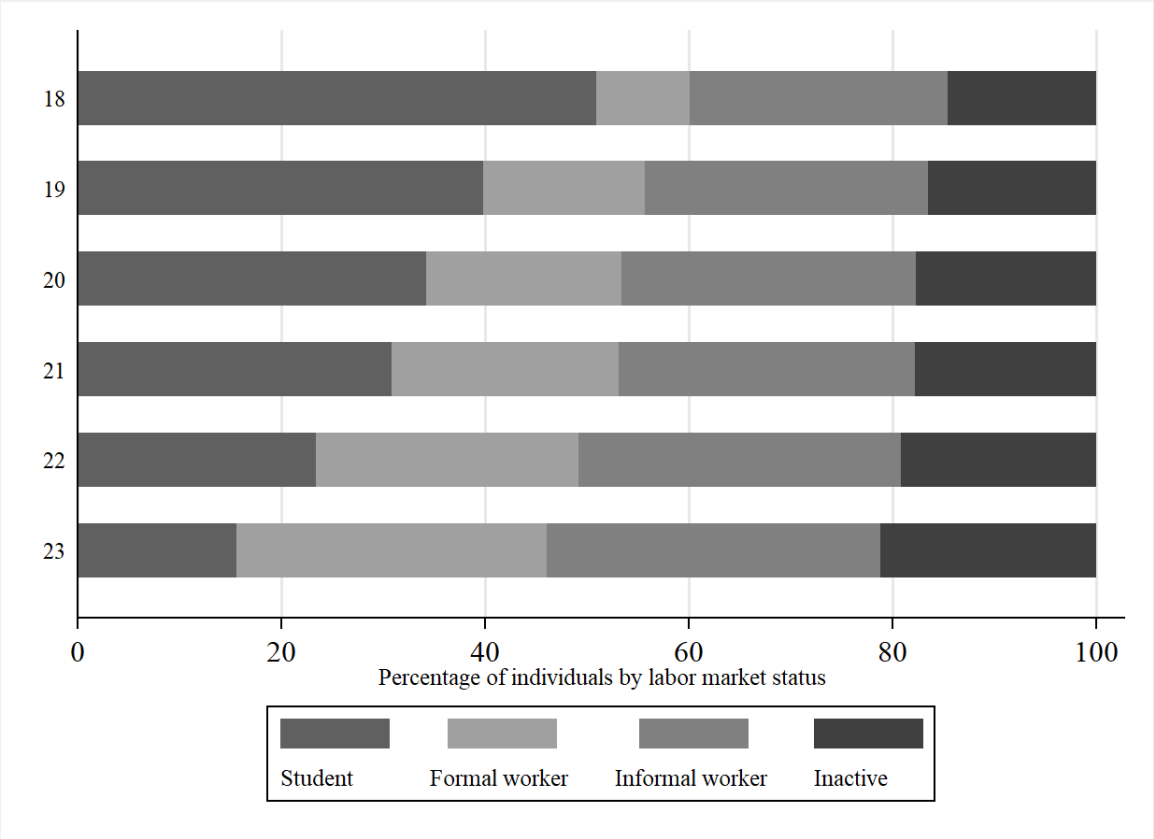
levels, such as the lack of schools or the opportunity cost of higher education. However, what is surprising is that high school enrollment is quite low even despite of its character of compulsory (since 2013). Furthermore, the enrollment rates of the last two years of college education and the first grades of graduate degrees are just 27 percent.

There is a substantial labor participation rate among individuals finishing upper secondary school. In Figure 2 I show the labor participation rate in Mexico for six age-cohorts of interest using the 2020 Mexican population census. Individuals who were born between 1997 and 1999 (21-23 years old in 2020) were not exposed to English instruction in primary school, while individuals who were born between 2000 and 2002 (28-20 years old in 2020) were potentially exposed. The decreasing participation rate in younger cohorts is consistent with the fact that younger kids have higher enrollment rates. Furthermore, from the 48 percent of 18 year old individuals not enrolled in school, 73 percent are employed. These data tell us that there is an important labor participation rate even for the youngest cohorts I will analyze in this paper, which suggests I will see them in the formal labor data even after considering that some will continue studying.

1.2 The National English Program in Mexico

The official language in Mexico is Spanish and, although there is no official estimation about the proportion of Mexicans who can speak and/or understand English, the reality is that just a few people have some kind of English knowledge in Mexico. For example, according

Figure 2: Composition of Mexican labor force by cohort (based on 2020 Mexican census data).



Note: Percentage of Mexicans in certain employment statuses (by cohort) are plotted. In the vertical axis I show age cohorts. For example, individuals who belong to the 2002 age cohort are 18 years old in 2020 and they have a labor participation rate of 33 percent, from which one seventh work in formal jobs. Almost 40 percent of the individuals in this cohort are still studying. The other extreme case is the 1997 age cohort where individuals are 23 years of age in 2020 and their labor participation rate is 54 percent, from which one third are working in the formal sector. Only 10 percent individuals in this oldest cohort are studying.

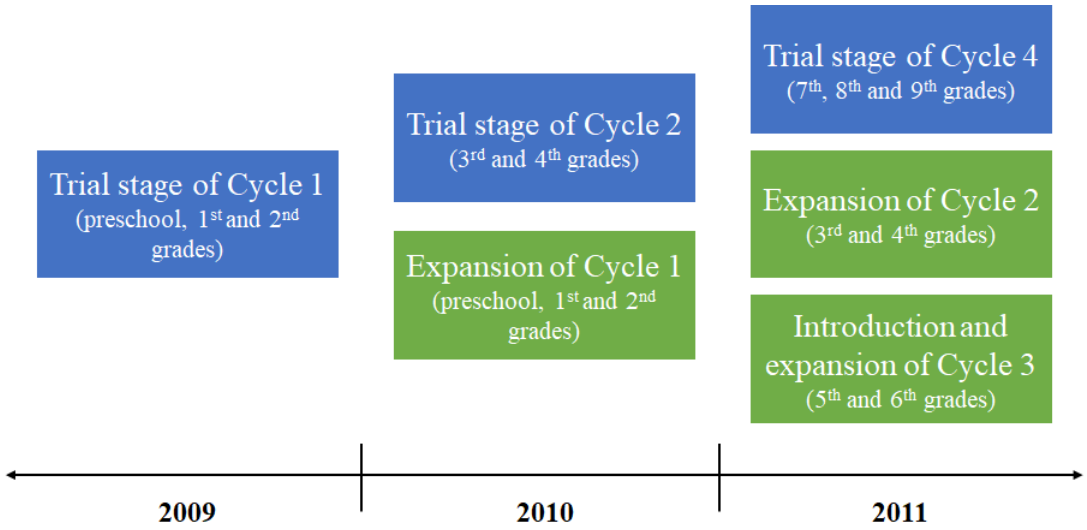
to the English Proficiency Index (IPE) generated by the company Education First, Mexico is classified as a country with low proficiency in English. Indeed, in 2020 this index ranked Mexico in the place 82 out of 100 non-English speaking countries all over the world, and in the place 18 out of 19 Latin American countries.[3] Furthermore, according to the survey of human capital in Mexico held by CIDAC (2008), four percent of the population declared to be able to read and listen in English, while only two percent of the population declared to know how to speak and write in English.

Knowing about the weakness among Mexicans to communicate in English, in 2009 the Mexican Ministry of Education (SEP, by its acronym in Spanish) launched a program called National English Program in Basic Education (NEPBE), which intended to introduce English instruction in all Mexican public elementary schools. Before this program, English instruction

[3]All these results are available in the 2020 edition of the EF English Proficiency Index report, published by Education First (2020).

was somewhat generalized among middle schools (because English instruction as a subject is compulsory in middle schools since 1926), but with the NEPBE English language education became compulsory since elementary school and it articulated the primary and secondary programs to give continuation of the English instruction between both educational levels.

Figure 3: The NEPBE implementation: trial and expansion stages



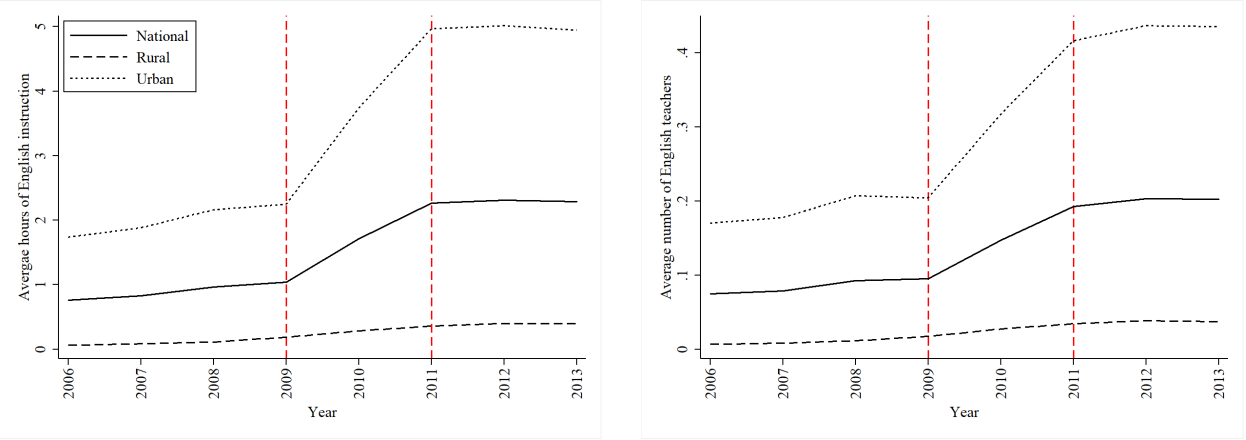
Note: The NEPBE was launched in 2009 as a trial stage with the so called Cycle 1. In 2010 the program continued the trial stage with Cycle 2 and expanded Cycle 1. Finally, in 2011 the program introduced and expanded Cycle 3, benefiting fifth and sixth graders.

Before the NEPBE, there were efforts to implement some kind of English instruction among elementary schools in 21 out of 32 Mexican states. In fact, in some of those states (Coahuila, Nuevo Leon, Sinaloa, Sonora and Tamaulipas) the already initiated English programs had strong fundamentals and had exceptional coverage compared to the other Mexican states (Sayer, 2015*a*,*c*; Ramírez-Romero and Sayer, 2016). However, the results were heterogeneous all around the country due to the differences in coverage, achievement levels, contents, English teachers supply and teaching hours. Hence, the NEPBE aimed to generalize the English instruction with the same content and with the same allocation of time. Nonetheless, the possibility of putting this program into practice has still been limited by the shortage of teachers trained for this purpose.

The limitations faced by previous state English programs were dealt by the central Mexican government through the implementation of the national English program by cycles and not by school grades, which improved continuity and articulation among the different grades and levels of the Mexican basic education system. The so called Cycle 1 comprises the third grade of preschool, as well as the first and second grades of elementary school; Cycle 2 includes third and fourth grades of elementary school; Cycle 3, comprises the fifth and sixth grades, while Cycle 4 includes all grades of middle school (SEP, 2011).

The NEPBE was launched in 2009 (as a pilot stage), however, the program really ex-

Figure 4: English instruction and English teachers over time (rural vs urban)



(a) Hours of English instruction **(b)** Number of English teachers

Source: Own elaboration with data from Mexican school census, Ministry of Public Education (SEP).

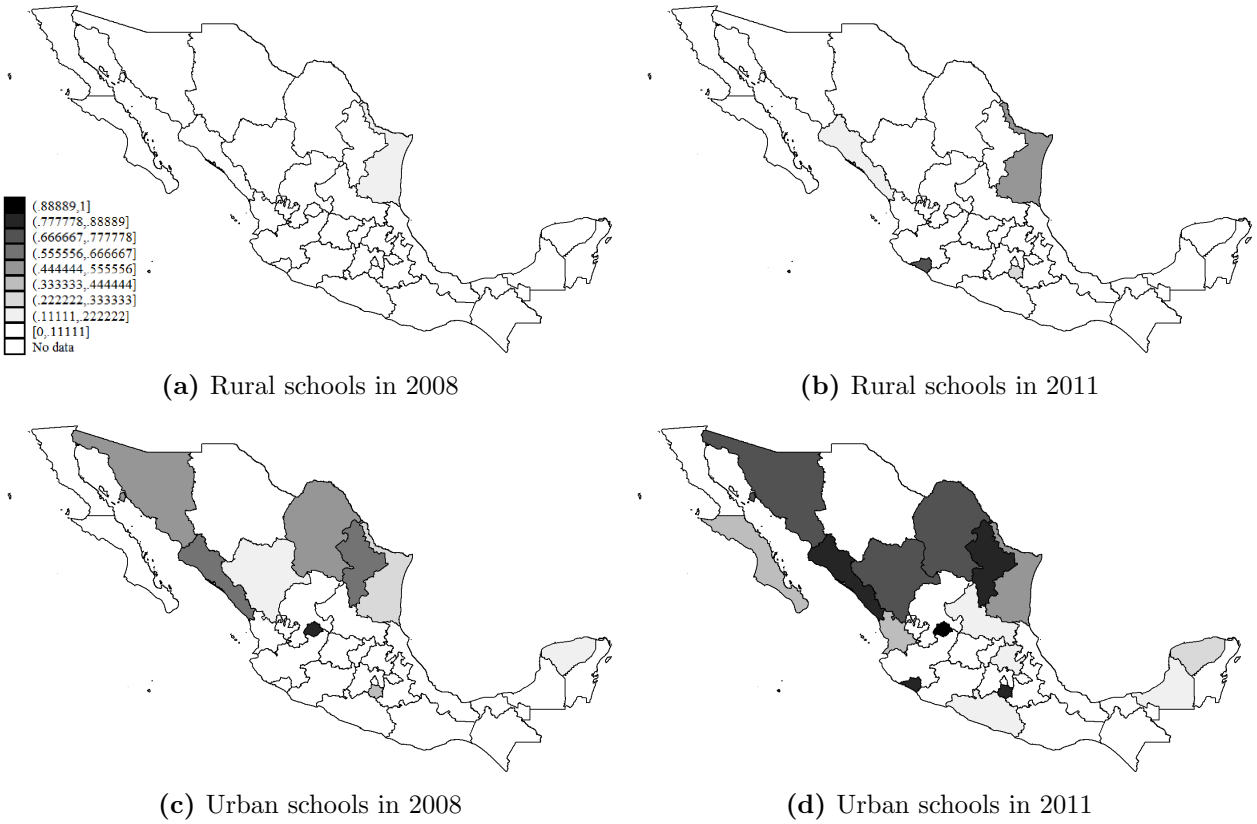
Note: Hours of English instruction are calculated dividing total hours in a school by total number of classes. Similarly, number of English teachers are calculated dividing total number of English teachers by total number of classes in a school. Vertical dotted line in 2009 highlights the implementation year of the NEPBE's trial stage, while the dotted line in 2011 highlights the expansion of the program.

panded in 2011 because before then the Mexican government suggested to implement it only among the first four grades of primary education (from first to fourth grades) and in a few randomly assigned schools. Hence, in 2011, fifth and sixth graders were exposed for the first time to the English program, but also more schools became beneficiaries (see Figure 3). Likewise, due to changes in the central government political party, in 2014 the NEPBE was reallocated to be part of a wider program called Program to Strength the Quality of Basic Education, playing a secondary role in the national agenda. Finally, in 2017 the program changed again to a national relevance and since then it is know as the National English Program (or PRONI, by its acronym in Spanish).

Most of the beneficiary schools saw a real change in hours of English instruction and in the number of English teachers until the year 2011. Furthermore, there was no significant change in rural areas. This latter was mainly due to the operation rules of the program, which prevented poor and marginalized schools to implement the English program since they do not have the equipment needed and/or well established commuting roads (as I explain below). On the other hand, it is more evident that the English program actually increased the hours of English instruction and the number of English teachers in 2011 (see Figure 4).

Indeed, there are two interesting features that characterize the implementation of the NEPBE. First, elementary schools in rural areas were less affected by the English program. Among the Mexican states, only Aguascalientes, Colima, Morelos, Sinaloa and Tamaulipas increased English instruction (between 2008 and 2011) in terms of the proportion of rural schools with some kind of English instruction (see panels (a) and (b) of Figure 5). On

Figure 5: Proportion of Mexican primary schools with English instruction (urban vs rural)



(a) Rural schools in 2008

(b) Rural schools in 2011

(c) Urban schools in 2008

(d) Urban schools in 2011

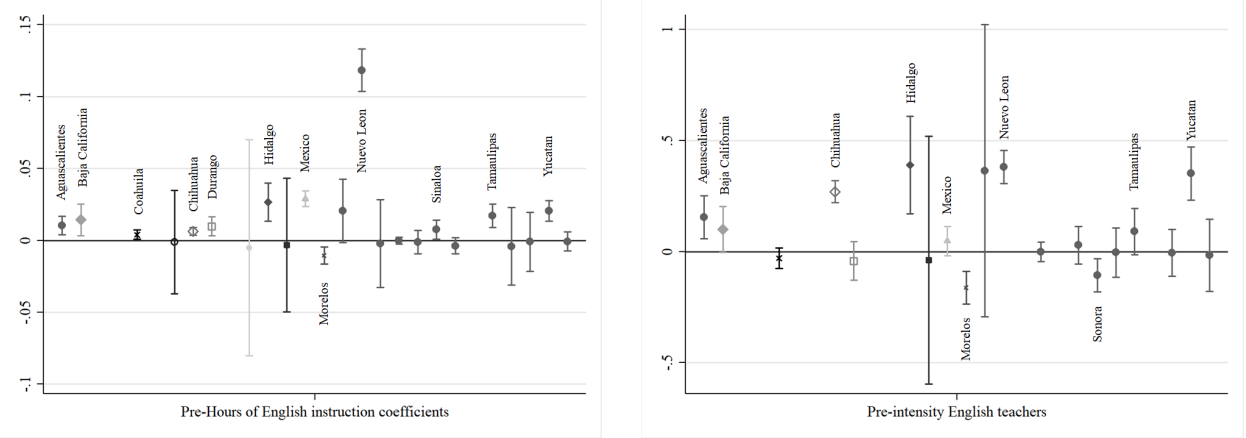
Note: Before the NEPBE, the English instruction was rare among rural schools. With the NEPBE, the proportion of rural schools with English instruction increased (between 2008 and 2011) in a few Mexican states: Colima, Morelos, Sinaloa and Tamaulipas. On the other hand, with the NEPBE, the proportion of urban schools with English instruction increased (between 2008 and 2011) in several northern (Baja California Sur, Coahuila, Colima, Durango, Nayarit, Nuevo Leon, San Luis Potosi, Sinaloa, Sonora and Tamaulipas) and some southern (Campeche, Colima, Guerrero, Hidalgo, Morelos and Yucatan) Mexican states.

the other hand, with the NEPBE the proportion of urban schools with English instruction increased (in the same period) in several northern (Baja California Sur, Coahuila, Colima, Durango, Nayarit, Nuevo Leon, San Luis Potosi, Sinaloa, Sonora and Tamaulipas) and some southern (Campeche, Colima, Guerrero, Hidalgo, Morelos and Yucatan) Mexican states (see panels (c) and (d) of Figure 5). This is consistent with the operation rules of the program.

Second, the relationship between pre-intensity and the change in the intensity of English instruction is positive for most of the Mexican states. This suggests that schools with less or no English exposure (before the English program) experienced a lower expansion of English instruction after the implementation of the NEPBE. In particular, this effect is substantial in the states of Nuevo Leon and Hidalgo (see panel (a) of Figure 6). Considering that the hours of English instruction could be a noisy variable because it is reported at school level, I also show analogous results where the dependent variable is the number of English teachers. The results are quite similar in terms of the sign and the significance of the coefficients (see

panel (b) of Figure 6).

Figure 6: Relationship between pre-intensity and change in intensity of English instruction



(a) Hours of English instruction

(b) Number of English teachers

Note: Coefficients of the pre-intensity of English instruction variable are plotted. These coefficients correspond to separate regressions per state, where the dependent variable is the change in the intensity of English instruction (panel (a)) or the number of English teachers (panel (b)).

1.3 Time allocation and beneficiary schools

A concern in this paper is that I do not observe variables that help me out to construct a human capital production function at individual or class level. However, I do know about the production function at school level in terms of the number of teachers and their schooling. Even though, it is not possible to know the allocation of time to each of the subjects. This latter is a limitation I solve by showing that the English program had no effect on other subjects in terms of test scores (see subsection 4.3 fora detailed discussion of these results).

The NEPBE did not require rules compliance from beneficiary schools other than to allocate the resources only to the purposes of the English program. For example, in terms of time allocation, we know that the English program implementation was flexible. The Mexican government offered guided sequences of contents that enabled teachers to carry out the adaptations demanded by specific instances of the Mexican education system. This flexibility was responsible for the way each school incorporated the English language subject into the regular curricula, in response to schools' heterogeneity. In this sense, the English program has worked mainly through the provision of economic resources and educational materials to Mexican schools seeking to implement English instruction as a subject. However, this program does not require neither rules compliance nor minimum expected outcomes.

Before the implementation of the NEPBE, schools already had a free allocation time known as "curricular autonomy", which they could use to teach the most convenient subjects

according to the students needs. With the implementation of the NEPBE some schools used part of their curricular autonomy to include English instruction. This means that the English program did not affect directly the minimum allocation of Spanish and Math instruction, but could have affected the support or bolster up of both subjects if a school used the curricular autonomy to help students with Math and Spanish.

Nowadays, the Mexican ministry of education suggests that the English language subject represents one eight (2.5 hours) of the effective teaching (20 hours). Likewise, the Spanish and Math instruction represents one half (10 hours) of the effective teaching time (one quarter or 5 hours each). Additionally, schools have 2.5 hours of curricular autonomy. Nonetheless, although the effective number of hours of Math and Spanish instruction has not changed since SEP launched the NEPBE, each school has the right and the freedom to structure or change the curricula according to the students needs. However, it is also true that, traditionally, all teachers in elementary schools cover at least one daily hour of Spanish and one of Math instruction, which is difficult to change even with the inclusion of the English subject in the regular curricula.

The English program's target population are preschools, primary and lower-secondary public schools. However, only schools that already had "English materials" for students, English teachers or a minimum of resources (such as computers, access to internet and easily accessible commuting roads) were eligible to be beneficiaries of the program. This is consistent with the positive coefficients plotted in Figure 6, which refer to the pre-intensity effect on change in hours of English instruction.

Hence, the national English program was initially implemented only in the state capitals or other larger urban areas. But even nowadays, in most of the states, the program coverage continues to focus on state capitals or major cities, neglecting smaller populations and marginalized areas. Furthermore, in many cases, the program has only served students enrolled in the morning shift or it coincides with the school's beneficiaries of the full time school (FTS) program.

According to the English program's rules, the central government transfers the economic resources to beneficiary schools through the school districts (which in some cases are managed by the state governments). School districts request the funds to the central government depending on the number of applications they received from the primary schools' principals. The applications usually contain standard formats where schools present an annual plan for the school year with objectives and strategies that are aligned to the central government goals. These applications should also include a plan for the implementation of English instruction in their regular curricula.

Currently, the Mexican government provides the NEPBE resources to public elementary schools that fulfill at least one of the following criteria:

- Being located in one of the 401 cities identified as part of the National Urban Sys-

tem that elaborates the Mexican population institute (CONAPO, for its acronym in Spanish).

- Being located in a county with more than 2,500 inhabitants.
- Being located in a community with paved access roads.
- Being located within two kilometers away from the corresponding central business district.
- Being located in a county considered as predominantly indigenous.
- Being located in one of the 121 communities classified as "magic villages" by the ministry of tourism.
- The school has to have the necessary facilities to implement the English program such as: a modern electric system, and access to computers with WiFi.
- School located in the Mexican southern states will be prioritized.

These criteria along with the evidence shown in Figure 3 and Figure 4 support the idea that the NEPBE did not benefit rural areas, which provides a source of variation that could be potentially exploited in future research.

1.4 Does English exposure imply the acquisition of English abilities?

The NEPBE increased the exposure to English language instruction through the incorporation of English language as a subject in the regular curricula of Mexican primary schools. The program increased the exposure to English instruction in (at least) six years, however, I am aware that this does not imply necessarily the formation of English language skills. Nevertheless, this English exposure could have had an effect increasing the familiarity with this foreign language.

The issue is that this is still an open question because there is no formal research that has studied the effect of the NEPBE on English language abilities. Furthermore, there is no measure of English knowledge in Mexico, which makes it more difficult to evaluate the effect of the program in terms of English language skills. In an initial try to "answer" this question, Székely, O'Donoghue and Pérez (2015) conducted a survey in Mexico as an instrument to measure English abilities and documented their findings in a report. The results shown in their report suggest that only 21 percent of the Mexican students survey showed some English knowledge and only three percent reached the expected standards by the Mexican government.

Nonetheless, the evidence points out that the results found by Székely, O'Donoghue and Pérez (2015) represent a lower bound measuring English abilities in Mexico. First, the timing; the authors conducted a survey in 2014 to middle school graduates, which means that these students had little or no exposure to English instruction in primary school because they were at this educational level during the trial stage of the program (see Figure 3 and Figure 4). Second, the sample; Székely, O'Donoghue and Pérez (2015) surveyed students without distinguishing those who studied in schools with English instruction and those who did not. This means that their sample contains both type of students, underestimating the proportion of individuals with English abilities. Finally, the geographical location; most students in their sample live in states with low proportion of primary schools with English instruction (Baja California, Jalisco, Guanajuato, Mexico, Puebla, Chiapas and Mexico City, with less than 11 percent of schools with English instruction) and only two states of the sample with a considerable high proportion (Nuevo Leon and Sinaloa, see Figure 5). These arguments are consistent with the findings of Sayer (2015*b*), who also suggests that not only the sample is erroneous to measure English ability post-NEPBE, but also the instrument was not appropriate and some of the questions were confusing to students.

In summary, we do not know whether the exposure to English instruction (through the NEPBE) can be translated into English language abilities, but what we do know is that the results are likely less negative as those reported by Székely, O'Donoghue and Pérez (2015).

2 Empirical strategy

The primary variables of interest in this paper are the labor market outcomes of individuals who were exposed to English instruction at early stages of life. In particular; the probability to work in the formal sector, wages, geographical mobility and occupational choices. In order to better understand the reduced form effect of English exposure on labor market it is important to understand first how English exposure affects student achievement. This is particularly useful in our context where we do not have measures of English language abilities. Indeed, understanding the relationship between English exposure and student achievement helps determining if English instruction is a substitute or complement of other abilities that contribute to the formation of human capital and, ultimately, to wages.

In my analysis, I use school by cohort variation in exposure to English instruction in Mexican primary schools to uncover the causal relationship between English instruction and labor market outcomes. Much of this variation is driven by the policy change induced by the implementation of the NEPBE.

The potential issue in this paper is the selection problem. Indeed, there could be differences between students who went to different schools, which are not necessarily related to differences in English language instruction. To deal with these issues I exploit the richness of my data set using a school fixed effects approach in both models I explain in this section.

Figure 7: Exposure to English instruction by cohort

Cohort	Primary school						Secondary school						College			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1997	2003	2004	2005	2006	2007	2008	2009	2010	**2011**	2012	2013	2014	2015	2016	2017	2018
1998	2004	2005	2006	2007	2008	2009	2010	**2011**	2012	2013	2014	2015	2016	2017	2018	2019
1999	2005	2006	2007	2008	2009	2010	**2011**	2012	2013	2014	2015	2016	2017	2018	2019	2020
2000	2006	2007	2008	2009	2010	**2011**	2012	2013	2014	2015	2016	2017	2018	2019	2020	
2001	2007	2008	2009	2010	**2011**	2012	2013	2014	2015	2016	2017	2018	2019	2020		
2002	2008	2009	2010	**2011**	2012	2013	2014	2015	2016	2017	2018	2019	2020			

Note: The rows in the figure represent the age cohorts and columns represent the school progression from primary to college (by grades). Cells market in dark gray suggest that those age cohorts were not exposed to English instruction in the indicated grades. Cells highlighted in light gray show the grades for which the cohorts were exposed to English instruction, according to the NEPBE (in bold the final year of expansion, 2011). Finally, white cells indicate grades in which traditionally any cohort has some kind of English instruction if they continue studying (upper secondary and college education).

2.1 Exposure to English instruction and labor market outcomes

Let us consider y_{isct} as the dependent variable, which can be any of the labor market outcomes of interest that characterize each individual i, who studied in school s, belongs to cohort $c = \{1997, 1998, 1999, 2000, 2001, 2002\}$ and whose labor market outcomes are observed in time $t = \{2018, 2019, 2020, 2021\}$. By the year 2021, all individuals in my sample had completed high school, which means that I am ruling out those students who decided to pursue a college or a graduate degree (as I explain below). However, this is not a concern because college and graduate enrollment rate is low in Mexico (about 27 percent in 2020, see Figure 1).

I concentrate my research on students who were exposed to English instruction during primary school due to the NEPBE. This means that students enrolled in fourth grade in 2011 (2002 cohort) had potentially five years of exposure in primary school and three more in lower secondary. Students enrolled in fifth grade in 2011 (2001 cohort) had three years of exposure in primary school (plus three years in lower secondary). Likewise, students enrolled in sixth grade in 2011 (2000 cohort) had exposure to only one year of English instruction in primary school (plus three in lower secondary). On the other hand, the three older age cohorts (1997-1999) were not exposed to the English instruction in primary school (see Figure 7). None of the 2000-2002 cohorts are expected to finish a college degree if they appear in the labor database between the years 2018 and 2021. However, the three oldest age cohorts (1997-1999) could have studied a college degree, so I restrict the age in the labor data to individuals who are 20 or younger in 2018, 21 or younger in 2019 and 22 or younger in 2020-2021, to avoid this potential issue.

In order to exploit the exposure to English variation (due to the program implementation rules), I construct an exposure to English instruction variable, $ExpEng_{sc}$, which takes into

account the variation of exposure by cohort and the variation in adoption among primary schools by averaging the number of hours of English instruction from first to sixth grade. Considering those two sources of variation: by school and by cohort, I propose the following school fixed effects equation:

$$y_{isct} = \alpha + \beta \cdot ExpEng_{sc} + \boldsymbol{X_{isct}\gamma} + \zeta_c + \nu_s + \tau_t + \varepsilon_{isct} \tag{1}$$

where I control for common cohort trends, ζ_c, unobserved attributes like quality of school, neighborhood effects and growing up environment, ν_s and time fixed effects, τ_t. I also include a set of control variables, $\boldsymbol{X_{isct}}$, which remove confounding effects of the exposure to English instruction on the dependent variable. For example, I control for test scores (when the worker was in primary school), gender, teacher's characteristics (in primary school), number of jobs, as well as age (as a proxy of experience) and its square as in the Mincer (1974) equation.

By including school FE I can deal with a lot of omitted variables problems, like those that are correlated to neighborhood and schools' characteristics, which are common characteristics of students within the same school. Hence, if we believe that the unobserved characteristics are constant over time, my empirical strategy will solve the selection problem and the estimator β will provide a causal interpretation of the effect of being exposed to English instruction at early stages of life on labor market outcomes.

2.2 Exposure to English instruction and student achievement

In this subsection I propose to look at the effect of exposure to English instruction on student achievement. This part of the analysis is important to better understand the mechanisms through which English instruction affects labor market outcomes. First, a natural mechanism could be a change in English language abilities, i.e., that the English instruction really had a positive effect developing English language skills among students exposed to the program. Hence, this new skill could have an effect on the labor market outcomes by expanding the possibility of finding well paid jobs or simply due to an "English premium" on wages. However, there are no measures of English ability in Mexico, which makes it difficult to test this mechanism (see subsection 1.3). Second, English instruction could affect other subjects (and, hence, other abilities), which could have implications on labor market outcomes. Indeed, adding time to English instruction could decrease the time allocated to language and/or Math subjects. Nonetheless, it could also be the case that English instruction strengthens language skills through a complementary between English and the native language, which is consistent with the existing research suggesting that taking foreign language classes increases language abilities.

Although I do not observe English abilities, I can argue that if English instruction does not have an effect on student achievement, the effect of English instruction on labor market outcomes is mainly through an improvement on English language skills. Indeed, as I discussed

above, the introduction of English instruction in primary schools could affect the teaching time of other subjects. If this is the case, labor market outcomes would be affected by the NEPBE because of either the rivalry in time among subjects or by the complementary on language abilities. Thus, I would find it difficult to separate the effect of English abilities and other cognitive abilities on earnings if the NEPBE had significant effects on language and Math subjects because these are the basic skills needed and used in the labor market.

To evaluate the effect of English instruction on students achievement I propose to use a similar specification as in Equation 1. The difference is that now the dependent variable is a measure of student achievement for different cohorts of sixth graders (as shown in Figure 7). The main independent variable of interest is the measure of exposure to English, $ExpEng_{sc}$, which I constructed exploiting the panel structure of my database as the average of the hours of English instruction over the six years that comprise primary school in Mexico (just as I explained above). The rest of the variables were previously explained in subsection 2.1.

Indeed, notice that using a regression of student achievement on English instruction would produce a biased estimator because of the selection problem. I solve this problem by proposing a school fixed effects regression, where I also fully control for time fixed effects (FE), τ_t. The selection bias is caused because beneficiary schools were not randomly assigned (after the trial stage). In this sense, we are concerned that better schools chose to adopt the NEPBE, i.e, schools with more information, better teachers or with more resources. Similarly, it could also be the case that these beneficiary schools are located in villages with wealthy neighborhoods and/or better access to services. To solve this problem, I fully control for school FE, ν_s, which capture school and regional differences (that are constant over time). Finally, I also control for students and schools' characteristics, $\boldsymbol{X_{isc}}$.

Under this model, the effect of English exposure on student achievement is captured by ϕ in the following equation:

$$test\_score_{isc} = \theta + \phi \cdot ExpEng_{sc} + \boldsymbol{X_{isc}}\boldsymbol{\gamma} + \zeta_c + \nu_s + \varepsilon_{isc} \tag{2}$$

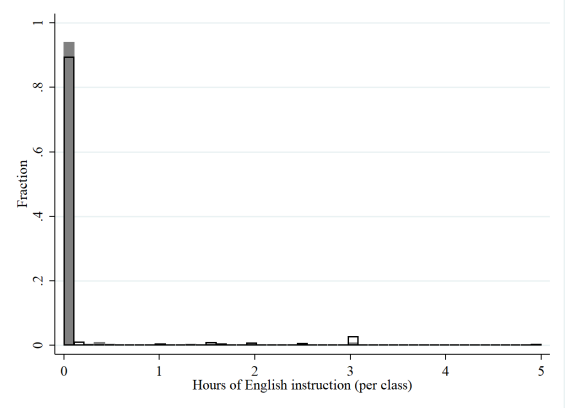
where $test\_score_{isc}$ is the sixth grade standardized test score of student i, who goes to school s and belongs to cohort c.

The potential selection bias caused by unobserved variables associated to the schools that adopted the English program is accounted by the inclusion of school FE. Indeed, it could be the case that schools with more resources are precisely the ones that had more information about the program and, thus, decided to participate. Under these conditions it is natural to think that also in these schools students will perform better in the standardized tests. However, all these differences among schools, which usually do not change in time, will be accounted by the school fixed effects.

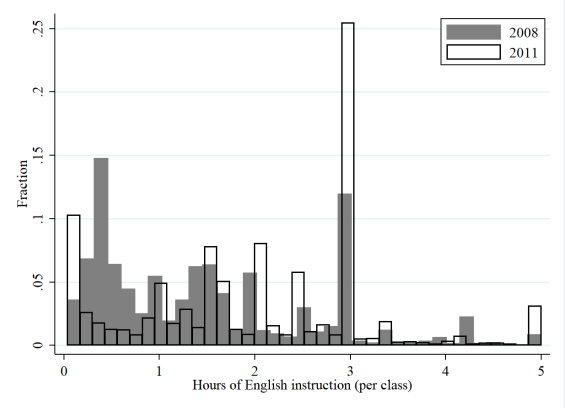
3 Data

I use three sources of data: school census, standardized test scores (in elementary schools) and formal labor market data.[4] I construct a novel database using these three sources of information because I am able to link each student to their primary schools and to their jobs when they go to the labor market. Indeed, I can observe each student in time since they are in third grade and then ten years later if they find a job in the formal sector. In this section I will explain how I use each data set.

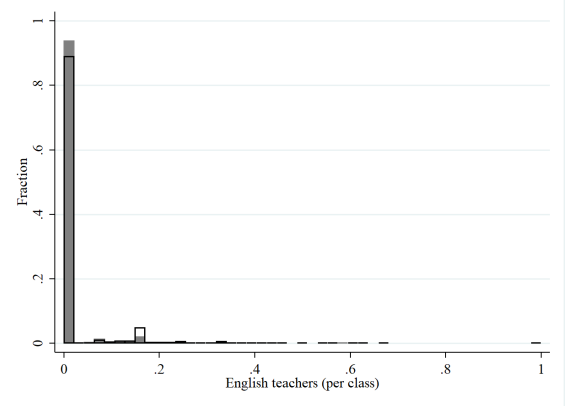
Figure 8: Hours of English instruction and English teachers distributions (2008 vs. 2011)



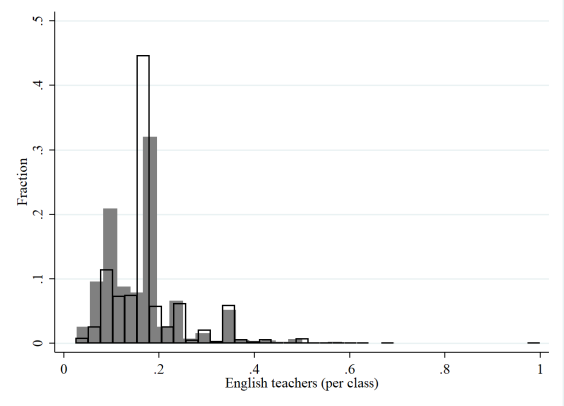
(a) Hours of English instruction



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



(c) Number of English teachers



(d) Number of English teachers (w/o zeros)

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.

[4] As I explain below, I have access to administrative records of more than 90 percent of all Mexican workers in the formal sector, but I do not observe those individuals who are inactive, students or that work in the informal sector.

3.1 Mexican school census

The first source of information I use in this research is the Mexican school census (also known as Statistics 911). The school census allows for identifying the schools that have offered English instruction before and after the implementation of the NEPBE. This is necessary because there are no official statistics about the list of beneficiary schools of this program, at least not before 2017. It is worth to mention that I exclude from the analysis those schools that are beneficiaries of the full time school (FTS) program for two reasons.[5] First, students' test scores could be positively affected by the FTS program. Second, the schools that participated in the FTS program were more likely to implement the English instruction in all grades (from first to sixth) and with more weekly hours. Furthermore, I only consider public elementary schools in the morning shift. Hence, removing all these aforementioned schools from my database let me obtain a cleaner effect of English instruction on labor market outcomes (and on school achievement).

Between the years 2008 and 2011, the distribution of weekly hours of English instruction and the number of English teachers shifted to the right, making evident the introduction of the NEPBE in Mexico. Indeed, as I explained in section 1, the trial stage of the NEPBE started in 2009 and progressively expanded from first to sixth grade. Thus, the English program reached out the planned expansion to all grades of primary school in 2011. In Figure 8, I compare the distribution of weekly hours and number of English teachers between 2008 and 2011 (i.e., one year before the implementation of the program and at the year of the final expansion).

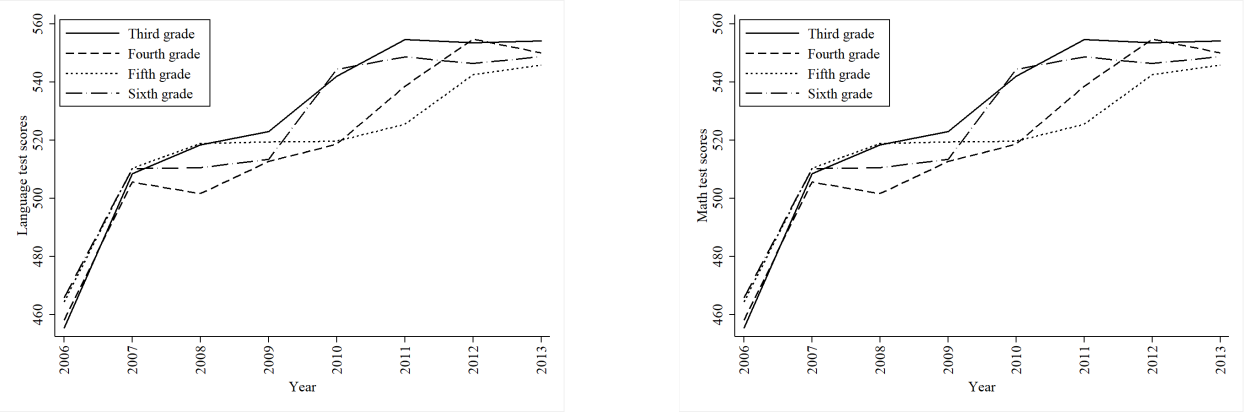
I measure weekly hours of English instruction as the ratio of total weekly hours of English instruction in each school and the total number of classes. For example, in a school where the reported weekly hours is 30, but the total number of classes is six (one section per grade), there are actually five weekly hours of English instruction per class (30/6 hrs.). In a similar example, another school also reports 30 hours of English instruction, but has two sections per grade. This latter school offers actually 2.5 weekly hours of English instruction per class (30/12 hrs.).

Most of the distribution of hours of English instruction concentrates at zero, suggesting that most of the Mexican schools do not offer English instruction. With the implementation of the program the density of zeros decreased (see panel (a) of Figure 8). In panel (b) of Figure 8 I show the distribution of weekly hours of English instruction before and after the implementation of the NEPBE, but without the zero values from the distribution. This figure suggests that some of the schools that had zero hours of English instruction in 2008 show a positive amount in 2011. Furthermore, many schools offered an amount of English instruction around the planned 2.5 hours (suggested by the Mexican government). And a considerable part of the distribution moved to three weekly hours.

[5]The full time school program was launched in 2007 with the objective of increasing the number of hours students spend at school. The trial phase of the program was implemented in 500 elementary and middle schools, located in 15 out of 32 Mexican states (Cabrera-Hernández, 2020).

English teachers are a scarce resource in Mexico. Hence, in many Mexican schools there is only one English teacher. I measure the number of English teachers as the ratio of teachers and classes. Following the examples of hours of English instruction, in a school with one teacher and six classes, the number of English teachers will be 0.16 (1/6), or 0.08 if the school has 12 classes (two sections per grade). In panel (d) of Figure 8, I show the distribution of English teachers between 2008 and 2011 without the zero values from the distribution. The distribution of English teachers moved to the right after the implementation of the NEPBE. Indeed, in 2008 the distribution concentrated between zero and 0.2 English teachers per class, on the other hand, with the English program the distribution concentrated between 0.18 and 0.38 English teachers per class.

Figure 9: Language and Math test scores over time and by grade



(a) Language test scores

(b) Math test scores

Note: Language and Math test scores are plotted over time and by grade. Two things can be noted from the graphs. First, test scores are increasing over time. Second, test scores of third graders are, on average, higher than fourth, fifth or sixth graders' test scores, every year.

3.2 ENLACE test scores

The second source of information is a standardized test known as ENLACE (National Evaluation of School Achievement in Educational Centers). ENLACE is a nation-wide standardized test that used to be applied to all students enrolled in public and private elementary and middle Mexican schools. This test was designed to examine student's Mathematics and Language (Spanish) achievement. This test was first applied in 2006 and discontinued in 2014, but replaced by the National Plan for the Learning Evaluation (PLANEA, for its acronym in Spanish).

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. This standardization takes into account that the test difficulty is different among grades and that it could change over time (as shown in Figure 9 test scores over time and by grade).

3.3 Social Security data

The third source of information corresponds to the labor data obtained from the Mexican Institute of Social Security (IMSS).[6] IMSS provides medical services and a contributory pension scheme to more than 90 percent of the formal workforce in Mexico (and nearby 40 percent of all Mexican workers).[7] This means that my estimates consider only individuals who work in the formal sector, which rises concerns for a second selection problem (selection into the sample). In section 4 I provide evidence of this problem, as well as a solution to obtain unbiased estimates. It is also worth mentioning that I am using the universe of Mexican students and almost the universe of formal workers (except for civil servants and the military), which provides more reliability of my proposed solution to the selection into the sample.

I use IMSS data from years 2018-2021, so there is enough time for individuals in my data set to enter into the labor market if they were exposed to the policy during their basic education. Indeed, we have already discussed that the NEPBE was first implemented in 2009 (as a pilot program) and expanded progressively with an important number of beneficiary schools in 2011. By the year 2011, three age cohorts were exposed to English language instruction: 2000, 2001 and 2002. In particular, students who were born in 2002 were exposed to a minimum of three years of English instruction in primary school, and potentially five years (see Figure 7). Next three older cohorts (1997-1999) were not exposed to English language instruction in primary school. Furthermore, none of these age cohorts could have finished college if I observe them in the labor market data and if I restrict age to individuals who are 22 or younger.

IMSS data is rich and complex. The data frequency is monthly, and each month could have more than one observation for the same worker because some workers have more than

[6]The data was accessed through the Econlab at Banco de México. The EconLab collected and processed the data as part of its effort to promote evidence-based research and foster ties between Banco de México's research staff and the academic community. Inquiries regarding the terms under which the data can be accessed should be directed to: econlab@banxico.org.mx.

[7]I estimate this percentages based on total IMSS affiliated workers (reported by IMSS itself) divided by the total formal employed workers (measured with the Mexican Labor Survey [ENOE]).

one job (in different or the same economic sector). To deal with this heterogeneity, I make some transformations to the data:

1. First, I take the average of the wages reported along one year, by worker, by economic sector and by employer.
2. Second, when a worker has multiple jobs, I drop the jobs with the lowest wages if those are non-permanent jobs.
3. Then, if there are individuals with permanent and non-permanent jobs, I use the information only of the permanent job.
4. Finally, for individuals who have more than one job with the same wage I choose the job in which they have worked most part of the year.

Wages reported in IMSS database are daily and before taxes. Furthermore, there is no information on the number of hours or days worked, but I assume that an employee works 30 days, on average. Hence, earnings reported in this paper correspond to a monthly wage (before taxes).

3.4 Economic industries

The effect of exposure to English instruction on wages can be better understood, as I mentioned before, by studying its first effect on student achievement. This analysis would increase our understanding of how taking a foreign language affects the formation of human capital. However, a second mechanism operates directly in the labor market. Indeed, if we are willing to believe that individuals actually acquired English abilities after the exposure to English instruction, there should be effects on occupational choices.

With the rapid growth of globalization among developing economies and the establishment of international companies in these economies looking for cheap labor force, it is tempting to argue that the formation of English abilities among some selected individuals could increase their likelihood to find more labor opportunities in export-oriented industries. This could be the case of maquiladoras in Mexico.

To study this mechanism I construct a variable that categorizes the economic industries reported in IMSS data using a similar method as the one proposed by Atkin (2016). First, I use the "correlation table" between the Law on General Import and Export Taxes (LIGIE) and the North American Industrial Classification System (NAICS), published by INEGI in 2021, to classify each of the products exported from Mexico in the NAICS system. Then, I took the ratio between the value of the exports in each NAICS industry and the GDP in that same industry (both expressed at a quarterly frequency and in current pesos).

I define an industry as export-oriented if more than 40 percent of the GDP in that industry is exported. The resulting export industries are: Air-Conditioning Equipment, Computer and Peripheral Equipment, Commercial and Service Industry Machinery, Communications Equipment, Electric Lighting Equipment, Forging and Stamping, Furniture, Hardware Manufacturing, Industrial Machinery, Magnetic and Optical Media, Measuring and Control Instruments, Medical Equipment and Supplies, Metalworking Machinery, Motor Vehicle, Other Electrical Equipment, Seafood Product and Packaging, and Ship and Boat Building.

The remaining industries such as: agriculture, utilities, construction and the other manufactures are classified as non-export industries. Likewise, I classify services in two main divisions: hospitality-telecommunications, and other services. I use these four categories to classify the 276 industry categories reported in the IMSS data. To accomplish this task I match the IMSS classification with the NAICS system using a match proposed by Banco de México (2021).

Table 1: Descriptive statistics

Variable	Mean	SD	Min	Max
Individual characteristics				
Female	0.39	0.49	0	1
Age	20.54	1.23	16	22
Language test score	-0.07	0.97	-2.84	3.49
Math test score	-0.05	0.97	-2.69	3.40
School characteristics				
Hours of English instruction	0.23	0.61	0	9.41
English teachers	0.02	0.05	0	1
Number of students (6th grade)	28.91	9.50	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
Labor market characteristics				
Wage (monthly pesos)	6,289	3,003	2,510	67,215
Permanent	0.81	0.39	0	1
Number of jobs (in a year)	1.46	0.81	1	17
Number of permanent jobs	1.18	0.82	0	14
Company size (workers)	1,913	5,200	1	92,972
Distance home-work (km)	107	264	0	2,029
Observations	3,588,166			

Note: These summary statistics correspond to the sample of students who were matched to their labor market outcomes. Furthermore, these statistics represent averages of all individuals in the sample (including all six cohorts) and also over the three years of labor data (2018-2021).

3.5 Descriptive statistics

The final data set, which contains the match between elementary school students and their labor market outcomes when employed in the formal sector, contains more than 3.5 million observations. This represents about one third of the total number of Mexican students in the six cohorts I study in this paper (1997-2002). The matched database includes individuals' characteristics, schools' characteristics and labor market variables such as wages, number of jobs in a year, number of permanent jobs and company size (see Table 1).

In the matched database, almost four of each 10 individuals are women and they are, on average, 20.5 years old. There are a lot of variation in terms of cognitive abilities (language and math), but the average individual is slightly to the left of the distribution of language and math test scores. The average school offers to each class about 14 minutes of English instruction per week. This last measure considers the zero hours of instruction offered by most of the Mexican primary schools, which explains the "short" average English instruction time. There are almost 29 students per class and more than 2/3 of the schools are located in urban localities.

Regarding labor market characteristics, the average worker earns a wage of 6,289 pesos per month, which is almost three times the poverty line in Mexico. Most of the workers in the formal sector (eight out of 10) have permanent jobs and the average worker has more than one job per year, which could be associated to the fact that I observe young workers (between 16 and 22 years of age) with a lot of mobility in their first years participating in the labor market. Finally, the average company size is 1,913 workers, but there is a lot of variation, going from one single worker to more than 92 thousand employees.

4 Results

In this section I show the results of the effect of English instruction on labor market outcomes and student achievement. The former are based on a school FE approach, where I am able to match students exposed to English language instruction to their labor market outcomes 10 years after finishing primary school. The later also exploits school FE, but I do not need to link individuals to their labor market outcomes because in this part of the analysis I am interested in studying their test scores at school. Using the school FE enables me to control for differences among schools that could generate a selection bias problem when estimating the effect of exposure to English instruction on labor market outcomes and student achievement.

Using the school FE strategy I solve the positive selection problem associated to the better characteristics of beneficiary schools over those that did not adopt the English program. This can be confirmed comparing columns 1-3 with column 4 of Table A.1 (from the Appendix), where the estimate is smaller once we mitigate the selection problem. I also offer a robustness check changing the measure of exposure to English instruction. Under this change, estimates

have the same direction and significance (see Panels C and D of Table A.1).

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
Panel A: Full sample				
Hrs English	-0.008***	-0.008***	-0.002	-0.001
	(0.001)	(0.001)	(0.005)	(0.001)
Observations	15,136,808	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.108	0.239	0.478	0.556
Panel B: Low enrollment sample				
Hrs English	-0.012	-0.001	-0.068	0.012
	(0.008)	(0.011)	(0.051)	(0.009)
Observations	1,383,627	231,142	231,142	231,142
Adjusted R^2	0.126	0.293	0.677	0.727
Panel C: Low enrollment sample (Men)				
Hrs English (β^M)	-0.018	-0.002	-0.126*	0.002
	(0.011)	(0.017)	(0.069)	(0.013)
Observations	667,238	148,036	148,036	148036
Adjusted R^2	0.151	0.296	0.680	0.729
Panel D: Low enrollment sample (Women)				
Hrs English (β^W)	-0.008	-0.010	0.029	0.025**
	(0.010)	(0.015)	(0.069)	(0.013)
Observations	716,389	83,106	83,106	83,106
Adjusted R^2	0.110	0.347	0.699	0.756
$\beta^M = \beta^W$ [p-value]	[0.016]	[0.542]	[0.257]	[0.306]
State of work FE	NO	YES	YES	YES

Note: This table shows the effect of the exposure to English instruction on labor market outcomes. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in the formal sector. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.1 English instruction and labor market outcomes

The exposure to English instruction at early stages of life could have changed the young students' human capital accumulation in several ways (which I discuss in the next subsection) and, hence, their labor market outcomes. If we compare students from the same schools, same ages and same backgrounds, but different exposure to English instruction, we could say something about the effect of this latter on labor market outcomes. The strategy I use to compare students with similar characteristics is to rely on a schools FE approach, which

I am able to do because of the richness of my database. My identifying assumption is that the unobservable characteristics that contribute to determine which schools are beneficiaries of the program from non-beneficiary schools, are fixed over time.

On the other hand, although I am able to match primary school students to their labor market outcomes, I can only observe students who decided to work in the formal sector when they participate in the labor market. This caveat causes an additional selection problem (selection into the sample). I show evidence of this selection problem in panel A, column 1 of Table 2. Indeed, my results suggest that one additional hour of English instruction reduces the probability of working in the formal sector by almost one percentage point. Later I show that this negative selection is explained by educational decisions of young individuals.

Due to this latter selection problem, we are worried that the estimates are downward bias, so we cannot interpret the exposure to English instruction on labor market outcomes as the real effect (see columns 2-4 of panel A). The intuition is that the potential high earners are being excluded from the formal labor sample. Thus, I only observe low skilled and low earners who would not had studied anyway, even after the exposure to English instruction.

To solve the selection into the (formal sector) sample I propose to use a sub-sample of individuals living in counties with low enrollment rates. From the previous analysis derived in Figure 2 of subsection 1.1, we know that if an individual is not working in the formal sector, he/she could be involved in one of three potential statuses: inactive, working in the informal sector or studying. Furthermore, we also noted that the proportion of individuals who are inactive and those working in the informal sector is pretty homogeneous among the age cohorts I study in this paper. However, the variation among different cohorts is potentially generated because the younger cohorts are still studying. Hence, since the selection into the sample is potentially caused by individuals who decide to pursue a high school or college degree, we could mitigate this selection problem by considering a sub-sample of individuals working in counties with low enrollment rates after finishing middle school.

I construct this sub-sample using the 2020 Mexican population census. In particular, I define a variable that identifies counties with low-enrollment rates in 2020 as follows:

1. I concentrate only on the youngest cohort (2002), which corresponds to the first year of high school when observed in 2020.
2. I identify the employment status of each individual: inactive, student or worker.
3. I create a variable for those individuals who are students, but not workers.
4. By county, I take the ratio of students to population (in the cohort 2002).
5. I create a variable that identifies counties with 38 or less percentage of individuals enrolled in high school.[8]

[8] I choose this percentage of enrollment based on a sensibility analysis that I summarize in Figure A.1.

Considering this criteria, I end up with a sub-sample of individuals with similar proportions of employment statuses among cohorts (see Figure A.2), which solves the selection into the sample problem. I provide evidence that this problem is solved in panel B, column 1 of Table 2, where I do not find any effect of English instruction on the probability of working in the formal sector.

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

	(1) Non-export industries	(2) Export industries	(3) Hospitality and Telecom	(4) Other Services
Panel A: Full sample				
Hrs English	-0.013***	0.000	0.013***	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	3,588,166	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.155	0.168	0.075	0.132
Panel B: Low enrollment sample				
Hrs English	0.007	-0.010	0.012**	-0.009
	(0.016)	(0.015)	(0.006)	(0.013)
Observations	231,142	231,142	231,142	231,142
Adjusted R^2	0.208	0.253	0.142	0.159
Panel C: Low enrollment sample (Men)				
Hrs English (β^M)	0.013	-0.009	0.011	-0.014
	(0.020)	(0.018)	(0.009)	(0.017)
Observations	148,036	148,036	148,036	148,036
Adjusted R^2	0.210	0.252	0.165	0.151
Panel D: Low enrollment sample (Women)				
Hrs English (β^W)	0.008	-0.008	0.014	-0.014
	(0.024)	(0.020)	(0.012)	(0.026)
Observations	83,106	83,106	83,106	83,106
Adjusted R^2	0.243	0.296	0.180	0.244
$\beta^M = \beta^W$ [p-value]	[0.072]	[0.000]	[0.822]	[0.727]
Shares	0.30	0.15	0.10	0.45

Note: This table shows the effect of the exposure to English instruction on economic industries. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in the formal sector. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Using this low-enrollment sample I find that the exposure to one additional hour of English instruction per week does not have any effect on wages. However, point estimates suggest that the effect could be negative with a more pronounced impact on women. Nevertheless, the estimated coefficients for men and women are not statistically different (see the t-test at the end of Table 2). On the other hand, although the low-enrollment sample helps to

overcome the selection into the sample, I find that in these low-enrollment counties men are less likely to work in the formal sector than women. This latter result suggests that there are more men enrolled in school than women in these type of counties. I study this potential implication below using data from the Mexican population census.

Table 4: Exposure to English instruction and company size (Social Security data)

	(1) Micro enterprise	(2) Small enterprise	(3) Medium enterprise	(4) Large enterprise
Panel A: Full sample				
Hrs English	-0.003***	-0.000	-0.004***	0.008***
	(0.001)	(0.001)	(0.002)	(0.002)
Observations	3,588,166	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.030	0.047	0.068	0.251
Panel B: Low enrollment sample				
Hrs English	-0.006	0.006	0.003	-0.003
	(0.007)	(0.009)	(0.011)	(0.012)
Observations	231,142	231,142	231,142	231,142
Adjusted R^2	0.053	0.082	0.105	0.307
Panel C: Low enrollment sample (Men)				
Hrs English (β^M)	-0.003	0.015*	-0.003	-0.010
	(0.007)	(0.009)	(0.014)	(0.016)
Observations	148,036	148,036	148,036	148,036
Adjusted R^2	0.068	0.093	0.115	0.295
Panel D: Low enrollment sample (Women)				
Hrs English (β^W)	-0.009	-0.005	0.017	-0.003
	(0.016)	(0.014)	(0.016)	(0.019)
Observations	83,106	83,106	83,106	83,106
Adjusted R^2	0.086	0.111	0.132	0.374
$\beta^M = \beta^W$ [p-value]	[0.499]	[0.959]	[0.327]	[0.658]
Shares	0.05	0.12	0.22	0.61

Note: This table shows the effect of the exposure to English instruction on the probability to work on certain types of company by size, measured by the number of workers according to the OECD (2022) classifications: Micro (workers$<$ 10), Small (10 $\leq$workers$<$ 50), Medium (50 $\leq$workers$<$ 250) and Large enterprise (workers$\geq$ 250). The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in the formal sector. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I also document that the exposure to English instruction has no effect on the geographical mobility (measured as the distance in miles from their home counties to their working counties) of Mexican workers. However, point estimates suggest that the exposure reduces

men's mobility while increases women's, although both estimates are not statistically different. Furthermore, I document that the exposure to English instruction increases women's likelihood of working on a state different from their home state, which suggests that women workers who were exposed to English instruction in primary school find more labor opportunities outside their home states. Men, on the other hand, are employed in places that are closer to their home counties within the same home state (see columns 3 and 4 of panels C and D in Table 2).

One of the mechanisms that explains the results shown previously is the type of jobs that the individuals find when they enter the labor market. There are two potential stories. First, that the exposure to English instruction increased the probability of working in jobs requiring English abilities, such as hospitality and telecommunications or, second, that this exposure increased the probability of working in companies that are mainly export-oriented. I provide evidence that the first story is more plausible (see Panel B of Table 3).[9] In fact, I show that workers who took English classes in primary school substitute jobs in export industries and other services for jobs in non-export industries, hospitality and telecommunications.

My results suggest that workers (that studied in low-enrollment counties) who were exposed to English instruction in primary school are more likely to work in hospitality and telecommunications. Furthermore, I find that men are more likely to work in non-export industries than women, and less likely to work in export industries (see Panels C and D of Table 3).

A finer analysis (shown in Table A.2 of the appendix) suggests that the exposure to English instruction increases the probability of working in construction, professionals and government. In particular, the effect on construction and professionals is driven by men, while I observe a reduction in the probability that women work in commerce.

In Table 4 I provide evidence that support the story of Table 3. In particular I show that men substitute jobs in micro, medium and large companies for jobs in small companies, which can be think of as international and export oriented enterprises. On the other hand, Women substitute jobs in small and large companies for jobs in medium ones.

[9]Notice that the estimates in panel A of are biased because there is a (negative) selection into the sample, plus individuals could choose different types of economic activities when working in the formal sector than in the informal sector. Indeed, I show evidence of this bias problem by comparing estimates of panel A and B.

Table 5: English instruction, education and earnings (Mexican census data)

	(1) Schooling	(2) Student	(3) Formal sector	(4) Informal sector	(5) Inactive	(6) ln(earnings) formal	(7) Labor force participation	(8) ln(earnings) all market
Panel A: Full sample								
Hrs English	-0.040	0.058***	-0.058***	-0.008	0.001	-0.049	-0.066***	-0.571***
	(0.066)	(0.008)	(0.008)	(0.005)	(0.004)	(0.044)	(0.007)	(0.059)
Observations	1,249,974	1,249,974	1,249,974	1,249,974	1,249,974	143868	1,249,974	1,249,974
Adjusted R^2	0.123	0.107	0.109	0.085	0.185	0.015	0.114	0.131
Panel B: Low-enrollment sample								
Hrs English	0.482*	0.017	0.017	-0.028	-0.017	0.198***	-0.011	0.252
	(0.246)	(0.015)	(0.019)	(0.024)	(0.020)	(0.075)	(0.034)	(0.398)
Observations	103,052	103,052	103,052	103,052	103,052	2,135	103,052	103,052
Adjusted R^2	0.110	0.052	0.049	0.282	0.449	0.076	0.300	0.198
Panel C: Low-enrollment sample (Men)								
Hrs English	0.523**	0.045**	0.037	-0.072**	-0.028***	0.165**	-0.036	0.469
	(0.257)	(0.019)	(0.030)	(0.031)	(0.010)	(0.073)	(0.048)	(0.703)
Observations	48,729	48,729	48,729	48,729	48,729	1,349	48,729	48,729
Adjusted R^2	0.087	0.056	0.066	0.071	0.026	0.083	0.078	0.135
Panel D: Low-enrollment sample (Women)								
Hrs English	0.373	-0.004	0.001	-0.000	-0.001	0.212	0.001	0.119
	(0.331)	(0.015)	(0.016)	(0.030)	(0.035)	(0.174)	(0.032)	(0.258)
Observations	54,323	54,323	54,323	54,323	54,323	786	54,323	54,323
Adjusted R^2	0.145	0.050	0.036	0.054	0.059	0.046	0.060	0.081

Note: This table shows the effect of the exposure to English instruction on education, economic statuses, and earnings. The sample contains all young Mexicans who belong to the age cohorts 1997-2002 and who are less than 22. All regressions include controls. In all these regressions the exposure variable is measured as the average hours of English instruction at county level. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 English instruction, education, earnings and occupations

The selection into the sample is potentially caused because some of the individuals I study are yet too young to participate in the labor market, so many of them are still enrolled in school (especially the youngest cohorts). Furthermore, it is likely that the individuals who were exposed to English instruction have a greater probability of continue studying compared to those who did not. Indeed, individuals who were exposed to English instruction have potentially more access to electronic and printed resources that are helpful at school, they open their set of options to pursue a higher degree and they ease their restrictions at understanding better the teaching material at school.

To prove this hypothesis, I propose to look at the schooling outcomes of the age cohorts I study in this paper, but in the 2020 Mexican population census. To perform this analysis, I create a variable of exposure to English instruction at the county level. This new exposure variable uses the previous one I created at the individual level (which had school by cohort variation) averaging the hours of English instruction among the primary schools located in the same county. Hence, the new exposure to English instruction variable has county by cohort variation. When creating this variable I assume that all individuals living in the same county have the same exposure to English instruction.

In Table 5 I offer evidence that individuals who were exposed to English instruction increase their probability of being enrolled at school. In particular, I find that one additional hour of English instruction increases the likelihood of continue studying by 6 percentage points (see column 2 of panel A). On the other hand, I do not find any significant effect on schooling, which is not surprising since individuals are too young to differ in years of schooling from cohort to cohort because some of them are still enrolled in school.

I also confirm my previous findings regarding the (negative) selection into the (formal sector) sample bias (see column 3 of panel A). Furthermore, I do not find any effect on the probability of working in the informal sector or the probability of being inactive. On the other hand, I also find a reduction in labor force participation by 7 percentage points, which is consistent with the increase in the probability of being enrolled at school. The effects on earnings are biased because of the selection problem.

Using the low enrollment sample (panel B of Table 5), I provide evidence of no-selection into the sample, so there is no effect on the probability of being enrolled at school nor on the years of schooling, but a substantial effect on earnings. Indeed, I find that an increase of one hour of English instruction (at county level) rises earnings by 20 percent among formal Mexican workers. However, although I find a positive point estimate for the effect of English instruction on earnings (of both, formal and informal workers), this estimate is not statistically significant (see column 8 of panel B). Among this individuals living in low enrollment counties, the exposure to English instruction increases schooling by almost half a year. This latter result suggests that English instruction also has positive effects on years of schooling, which could be proven in future research once individuals be older enough to rule

out the possibility that they are still enrolled in school.

In panel C of Table 5 I find that the exposure to English instruction increases men's likelihood of being enrolled in school as well as their years of schooling. Furthermore, there is an additional effect reducing the probability of working in the informal sector or being inactive. However, there is no effect on the probability of working in the formal sector. Hence, since there is no selection problem, I find that exposure to one additional hour of English instruction in primary school increases men's earnings by 17 percent. Consistent with my results using the social security data, I do not find any significant effect on women's earnings.

Table 6: Exposure to English instruction and occupations (Mexican census data)

	(1) Agri-culture	(2) Com-merce	(3) Services	(4) Crafts	(5) Manual Work	(6) Skilled Job	(7) Pro-Tech
Panel A: Full sample							
Hrs English	-0.023***	0.035**	0.035**	0.007	0.005	-0.005	-0.055***
	(0.004)	(0.011)	(0.011)	(0.008)	(0.009)	(0.008)	(0.015)
Observations	548,611	548,611	548,611	548,611	548,611	548,611	548,611
Adjusted R^2	0.295	0.042	0.022	0.058	0.048	0.083	0.163
Panel B: Low-enrollment sample							
Hrs English	0.049	-0.008	-0.039**	-0.174***	0.063**	0.117***	-0.008
	(0.067)	(0.023)	(0.015)	(0.023)	(0.028)	(0.016)	(0.045)
Observations	42,972	42,972	42,972	42,972	42,972	42,972	42,972
Adjusted R^2	0.249	0.078	0.064	0.082	0.042	0.061	0.095
Panel C: Low-enrollment sample (Men)							
Hrs English	0.057	0.027	0.009	-0.168***	-0.011	0.138***	-0.052
	(0.076)	(0.018)	(0.009)	(0.037)	(0.020)	(0.020)	(0.041)
Observations	32,713	32,713	32,713	32,713	32,713	32,713	32,713
Adjusted R^2	0.172	0.039	0.072	0.066	0.042	0.053	0.050
Panel D: Low-enrollment sample (Women)							
Hrs English	0.005	-0.136***	-0.236***	-0.156***	0.367***	0.036	0.120**
	(0.056)	(0.051)	(0.078)	(0.034)	(0.094)	(0.032)	(0.048)
Observations	10,259	10,259	10,259	10,259	10,259	10,259	10,259
Adjusted R^2	0.238	0.085	0.061	0.338	0.125	0.248	0.169
Shares	0.09	0.16	0.11	0.11	0.18	0.13	0.22

Note: This table shows the effect of the exposure to English instruction on occupations. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in some economic activity. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

One of the mechanisms that could explain the effect on wages and the gender differences is the occupational choices of individuals who were exposed to English instruction. These

decisions could be biased if we study all individuals who self-selected to drop-out school to participate in the labor market. Thus we could be excluding potentially high ability and high earner individuals (as I showed in column 3, panel A of Table 5). To mitigate the selection problem, I consider again the sub-sample of individuals living in counties with low (high school and college) enrollment rates. I find that exposure to English instruction reduces the probability of working in occupations offering a service and crafts occupations, while increases the probability of working on occupations requiring manual (skilled and unskilled) work (see panel B of Table 6). This latter result could be associated to the fact that these individuals are finding jobs at international companies manufacturing in Mexico, such as automotive and other maquiladoras. Interestingly, I find that only men increase their probability of working on skilled jobs, while women does only on unskilled jobs (manual work, see panels C and D of Table 6). This could explain the rise on earnings among men but not women.

4.3 English instruction and student achievement

Results from estimating Equation 2 provide evidence that the English program had no effect on test scores. Indeed, I find that the NEPBE had no effect on neither Language nor Math test scores. As I explained in subsection 2.2, school FE in this model account for most of the selection bias problem. However, I will also explore the possibility of providing more robust results by using a different identification strategy. In particular, I could use the exogenous variation generated by the English program as an instrument of English instruction to remove any remaining selection bias.

I show the results of estimating a school FE model exploiting the panel structure of my database in Table 7. To obtain the results of this table, I am using the same age cohorts I examined with the labor market outcomes model, so we can think about this as analyzing the Spanish and Math skills at the end of primary school due to the English exposure for those same individuals I examine their labor market outcomes.

There are three interesting things to notice from this table.

1. There is selection into which schools offer English instruction. When I control for school FE, taking care of many time invariant characteristics of schools and neighborhoods, the estimated coefficient associated to the exposure to English instruction variable goes down.

2. I find no effect of English instruction on test scores. This means that English instruction did not reduce skills in Spanish and Math, as feared if more time is devoted to English at the expense of other subjects. On the other hand, it did no increase skills, suggesting no complementary between English and Spanish.

3. The previous finding of no effect on Math and Spanish suggests that when I estimate the effects of English instruction on labor market outcomes, it is likely not be driven

by learning in other subjects, so it could be primarily interpreted as the direct effect of the acquisition of English language skills.

Table 7: Exposure to English instruction and student achievement

	(1) Language 6th	(2) Language 6th	(3) Math 6th	(4) Math 6th
Panel A: Full sample in Social Security data				
Hrs English	0.0315***	-0.0141*	0.0153***	-0.0169*
	(0.0034)	(0.0085)	(0.0037)	(0.0096)
Observations	3,588,166	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.397	0.450	0.404	0.465
Panel B: Low enrollment sample				
Hrs English	0.0463	0.0388	0.0099	0.0052
	(0.0449)	(0.0839)	(0.0338)	(0.0682)
Observations	205,116	205,116	205,116	205,116
Adjusted R^2	0.345	0.443	0.375	0.477
Panel C: Low enrollment sample (Men)				
Hrs English (β^M)	0.0599	0.0566	0.0156	-0.0056
	(0.0508)	(0.0931)	(0.0372)	(0.0836)
Observations	131,359	131,359	131,359	131,359
Adjusted R^2	0.306	0.426	0.364	0.481
Panel D: Low enrollment sample (Women)				
Hrs English (β^W)	0.0188	-0.0123	-0.0026	0.0208
	(0.0374)	(0.0915)	(0.0334)	(0.0725)
Observations	73,757	73,757	73,757	73,757
Adjusted R^2	0.363	0.484	0.389	0.518
$\beta^M = \beta^W$ [p-value]	[0.3600]	[0.4778]	[0.8484]	[0.8916]
State FE	YES	NO	YES	NO
School FE	NO	YES	NO	YES

Note: This table shows the effect of exposure to English instruction on test scores. All regressions include controls and school FE. The sample contains students who later in life will work in the formal sector and who studied primary school in counties where the upper-secondary and college enrollment rates are low. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Conclusions and discussion

In this paper I evaluate the effect of being exposed to English instruction at early stages of life on labor market outcomes, in the context of a non-English speaking country. For this purpose, I construct a novel database in which I observe primary school students in all Mexican schools, their student achievement and their labor market outcomes 10 years after

graduation from primary school. I exploit the school-cohort variation of exposure to English language instruction in a school FE model, which allows me to mitigate the selection problem by controlling for variables that are correlated to neighborhood and schools' characteristics, which are common features of students within the same school.

Indeed, using a school fixed effects approach, I find that exposure to English instruction reduces the likelihood that an individual participates in formal sector employment. It is likely that this result is due to exposure to English instruction affecting enrolling in high school and college, as my analysis focuses on young adults aged 16-22 (the recency of the NEPBE means the affected cohorts are still young). In fact, I provide evidence that exposure to English instruction increases the probability of being enrolled in school.

Focusing on a sub-sample that is unlikely to be enrolled by age 16, I find that men substitute jobs in non-export industries for jobs in export-industries after being exposed to English instruction, while women also substitute for jobs in hospitality and telecommunications. I also document that the exposure to English instruction increases wages for men but not for women. This gender difference could be explained because women substitute jobs in large companies for jobs in small and medium ones. Furthermore, I find that exposure to English instruction reduces the probability of working in occupations involving services and crafts occupations, while increases the probability of working on occupations requiring manual (skilled and unskilled) work. This latter result could be associated to the fact that these individuals are finding jobs at international companies manufacturing in Mexico, such as automotive and other maquiladoras. Interestingly, I find that after exposure to English instruction men increase their probability of working on skilled jobs, while women do only on unskilled jobs (manual work).

I propose to explore the role of the two other potential mechanisms through which exposure to English language instruction affects labor market outcomes. First, through English language skills. Indeed, it is natural to believe that individuals will develop some kind of English abilities after being exposed to English instruction. Second, through other related cognitive abilities. Here I identify two potential effects: 1) a positive effect because a potential complementarity between English and Spanish, and 2) a negative effect as a consequence of rivalry among subjects due to a change in teaching time allocation.

I do not have a measure of English abilities, however, I am able to explore the role of the second mechanism by looking at student achievement. My results suggests that there is selection into which schools offer English instruction, confirming the selection problem we were concerned. Furthermore, I find no effects on Language and Math test scores, which suggests that the estimated effect of exposure to English language on wages is not reflecting changes to general cognitive skills. These findings are consistent with exposure to English instruction raising wages through increasing English language skills.

References

Angrist, Joshua, Aimee Chin, and Ricardo Godoy. 2008. "Is Spanish-only schooling responsible for the Puerto Rican language gap?" *Journal of Development Economics*, 85(1-2): 105–128.

Angrist, Joshua D, and Victor Lavy. 1997. "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco." *Journal of Labor Economics*, 15(1, Part 2): S48–S76.

Atkin, David. 2016. "Endogenous skill acquisition and export manufacturing in Mexico." *American Economic Review*, 106(8): 2046–85.

Azam, Mehtabul, Aimee Chin, and Nishith Prakash. 2013. "The returns to English-language skills in India." *Economic Development and Cultural Change*, 61(2): 335–367.

Banco de México, BM. 2021. "Correspondencia de actividades dentro de la base de micro datos del IMSS con el Sistema de Clasificación Industrial de América del Norte 2018." Mexico City, Mexico, Dirección de Análisis Sobre Precios, Economía Regional e Información, Banco de México.

Bleakley, Hoyt, and Aimee Chin. 2004. "Language skills and earnings: Evidence from childhood immigrants." *Review of Economics and statistics*, 86(2): 481–496.

Bleakley, Hoyt, and Aimee Chin. 2010. "Age at arrival, English proficiency, and social assimilation among US immigrants." *American Economic Journal: Applied Economics*, 2(1): 165–92.

Borjas, George J, and Rachel M Friedberg. 2009. "Recent trends in the earnings of new immigrants to the United States." *NBER working paper series*, w15406: 1–56.

Cabrera-Hernández, Francisco. 2020. "Does Lengthening the School Day Increase School Value-Added? Evidence from a Mid-Income Country." *The Journal of Development Studies*, 56(2): 314–335.

Chakraborty, Tanika, and Shilpi Kapur Bakshi. 2016. "English language premium: Evidence from a policy experiment in India." *Economics of Education Review*, 50: 1–16.

Charles-Leija, Humberto, and Aldo Josafat Torres. 2022. "English Proficiency and Salary in Mexico." *Analisis Económico*, 30(94): 167–180.

Chiswick, Barry R. 1991. "Speaking, reading, and earnings among low-skilled immigrants." *Journal of labor economics*, 9(2): 149–170.

Chiswick, Barry R, and Paul W Miller. 1995. "The endogeneity between language and earnings: International analyses." *Journal of labor economics*, 13(2): 246–288.

Chiswick, Barry R., and Paul W. Miller. 2015. "International Migration and the Economics of Language." In *Handbook of the Economics of International Migration*. Vol. 1 of *Handbook of the Economics of International Migration*, , ed. Barry R. Chiswick and Paul W. Miller, 211–269. North-Holland.

CIDAC. 2008. *CIDAC survey about human capital in Mexico.* Mexico City:Centro de Investigacion para el Desarrollo, A.C.

Delgado Helleseter, Miguel Antonio. 2020. "English Skills and Wages in a Non-English Speaking Country: Findings from Online Advertisements in Mexico." *The International Journal of Interdisciplinary Global Studies*, 15(3): 13–27.

Dustmann, Christian. 1994. "Speaking fluency, writing fluency and earnings of migrants." *Journal of Population economics*, 7(2): 133–156.

Dustmann, Christian, and Arthur van Soest. 2001. "Language fluency and earnings: Estimation with misclassified language indicators." *Review of Economics and Statistics*, 83(4): 663–674.

Education First, EF. 2020. *EF English Proficiency Index: A Ranking of 100 Countries and Regions by English Skills.* United States:Education First.

Eriksson, Katherine. 2014. "Does the language of instruction in primary school affect later labour market outcomes? Evidence from South Africa." *Economic History of Developing Regions*, 29(2): 311–335.

Grenier, Gilles. 1984. "The effects of language characteristics on the wages of Hispanic-American males." *Journal of Human Resources*, 35–52.

Hayfron, John E. 2001. "Language training, language proficiency and earnings of immigrants in Norway." *Applied Economics*, 33(15): 1971–1979.

Lang, Kevin, and Erez Siniver. 2009. "The return to English in a non-English speaking country: Russian immigrants and native Israelis in Israel." *The BE Journal of Economic Analysis & Policy*, 9(1).

McManus, Walter, William Gould, and Finis Welch. 1983. "Earnings of Hispanic men: The role of English language proficiency." *Journal of Labor Economics*, 1(2): 101–130.

Mincer, Jacob. 1974. "Schooling, earnings and experience." *National Bureau of Economic Research*.

Ramírez-Romero, José Luis, and Peter Sayer. 2016. "The teaching of English in public primary schools in Mexico: More heat than light?" *education policy analysis archives*, 24: 84.

Rivera-Batiz, Francisco L. 1990. "English language proficiency and the economic progress of immigrants." *Economics Letters*, 34(3): 295–300.

Sayer, Peter. 2015*a*. "Expanding global language education in public primary schools: The national English programme in Mexico." *Language, Culture and Curriculum*, 28(3): 257–275.

Sayer, Peter. 2015*b*. "A response to the Sorry report on the state of the PNIEB and English language education in Mexican public schools." *Mextesol Journal*, 39(4): 1–11.

Sayer, Peter. 2015*c*. ""More & earlier": Neoliberalism and primary English education in Mexican public schools." *L2 Journal*, 7(3).

SEP. 2011. *National English Program in Basic Education. Second Language: English.* Mexico:Secretaria de Educacion Publica.

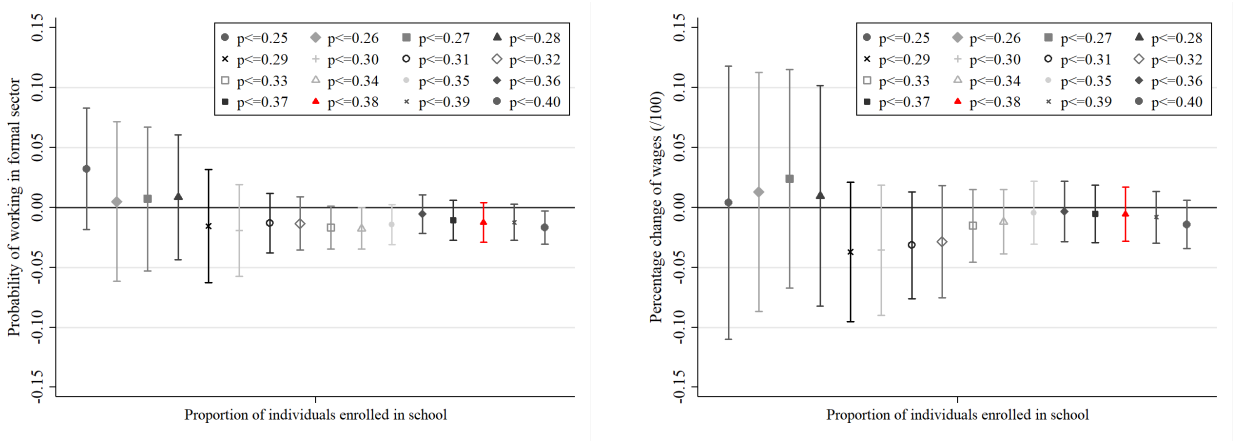
Székely, M, JL O'Donoghue, and H Pérez. 2015. "El estado del aprendizaje del inglés en México." In *Sorry: el aprendizaje del inglés en México.* , ed. Jennifer L. O´Donoghue, 83–97. Mexico:Mexicanos Primero.

Taylor, Stephen, and Marisa von Fintel. 2016. "Estimating the impact of language of instruction in South African primary schools: A fixed effects approach." *Economics of Education Review*, 50: 75–89.

Williams, Donald R. 2011. "Multiple language usage and earnings in Western Europe." *International Journal of Manpower.*

Appendix

Figure A.1: Effect of exposure to English instruction on wages (by proportion of enrollment)



(a) Formal sector **(b)** Earnings

Source: Own elaboration with data from Mexican Social Security data.

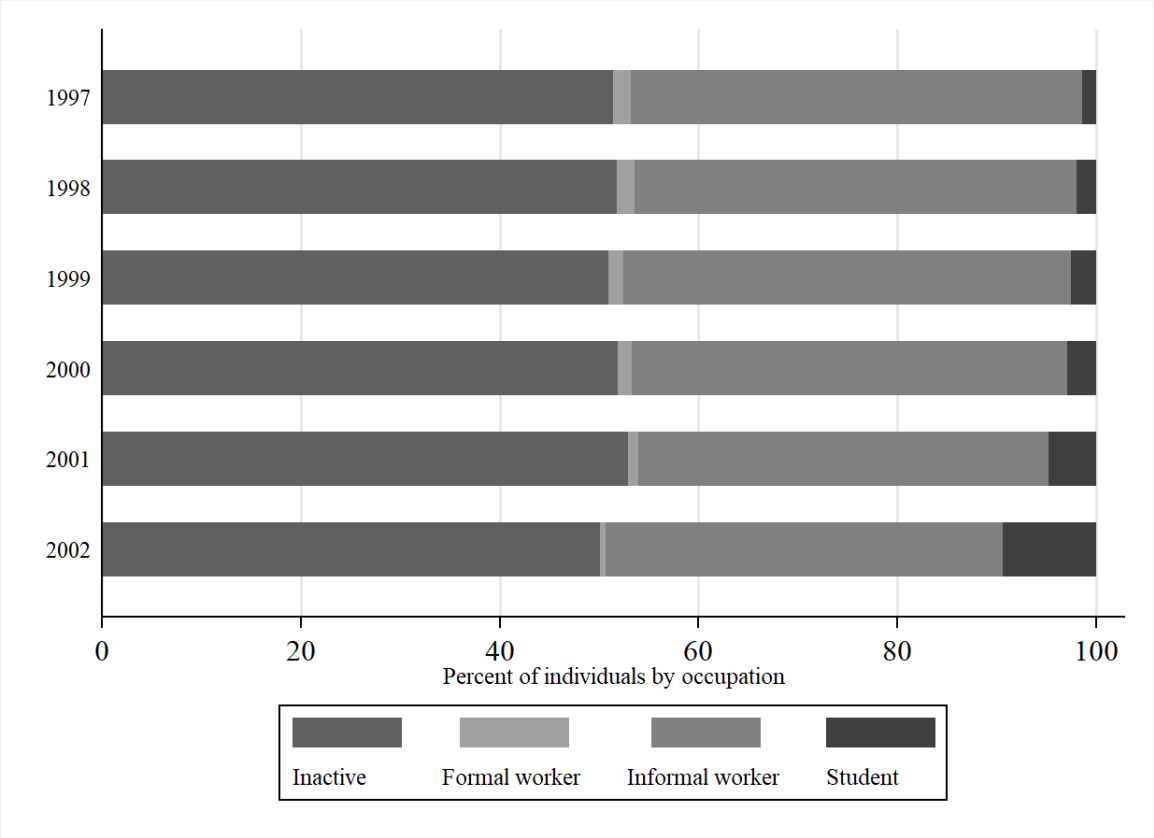
Note: Hours of English instruction are calculated as the average hours of English instruction by county. Estimates from different regressions are plotted, where the difference comes from the samples used. Estimates where the proportion of individuals enrolled in school is less than 9, 10 and 11 percent are unbiased.

Table A.1: English instruction and the selection problem (Social Security data)

	(1) ln(wage)	(2) ln(wage)	(3) ln(wage)	(4) ln(wage)
Panel A: Hours of English instruction				
Hrs English	-0.006***	-0.009***	-0.008***	-0.014***
	(0.002)	(0.001)	(0.001)	(0.002)
Observations	3,588,166	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.192	0.219	0.235	0.238
Panel B: Hours of English instruction (low enrollment)				
Hrs English	-0.000	-0.007	-0.004	-0.001
	(0.008)	(0.007)	(0.008)	(0.011)
Observations	231,142	231,142	231,142	231,142
Adjusted R^2	0.211	0.247	0.297	0.293
Panel C: Number of English teachers				
Eng Teachers	-0.074***	-0.116***	-0.101***	-0.145***
	(0.025)	(0.014)	(0.015)	(0.029)
Observations	3,588,166	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.192	0.219	0.235	0.238
Panel D: Number of English teachers (low enrollment)				
Eng Teachers	-0.012	-0.139	-0.113	-0.167
	(0.120)	(0.092)	(0.100)	(0.185)
Observations	231,142	231,142	231,142	231,142
Adjusted R^2	0.211	0.247	0.297	0.293
State FE	YES	NO	NO	NO
County FE	NO	YES	NO	NO
Locality FE	NO	NO	YES	NO
School FE	NO	NO	NO	YES

Note: This table shows the effect of the exposure to English instruction on wages under different fixed effects levels to see how the selection problem is solved using school FE. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in the formal sector. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.2: Composition of Mexican labor force by cohort (low-enrollment sample).



Note: Percentage of Mexicans in certain employment statuses (by cohort) are plotted. Notice that, in this sub-sample, the proportion of individuals performing certain occupations is relatively homogeneous across cohorts.

Table A.2: Exposure to English instruction and economic industries (Social Security data)

	(1) Agri-culture	(2) Con-struction	(3) Manu-facture	(4) Com-merce	(5) Profes-sionals	(6) Govern-ment	(7) Hospi-tality
Panel A: Full sample							
Hrs English	-0.005***	-0.010***	-0.003	0.013***	-0.007***	-0.001**	0.013***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Observations	3,588,166	3,588,166	3,588,166	3,588,166	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.291	0.119	0.194	0.072	0.061	0.025	0.075
Panel B: Low enrollment sample							
Hrs English	-0.013**	-0.028**	0.030*	0.006	-0.010	0.002	0.013**
	(0.006)	(0.012)	(0.017)	(0.011)	(0.010)	(0.004)	(0.006)
Observations	231,264	231,264	231,264	231,264	231,264	231,264	231,264
Adjusted R^2	0.321	0.307	0.272	0.135	0.065	0.049	0.142
Panel C: Low enrollment sample (Men)							
Hrs English (β^M)	-0.006	-0.027	0.033	-0.003	-0.011	0.004	0.011
	(0.010)	(0.017)	(0.022)	(0.015)	(0.010)	(0.005)	(0.009)
Observations	148,095	148,095	148,095	148,095	148,095	148,095	148,095
Adjusted R^2	0.341	0.335	0.281	0.141	0.068	0.065	0.165
Panel D: Low enrollment sample (Women)							
Hrs English (β^W)	-0.023***	-0.016	0.028	0.008	-0.011	-0.001	0.014
	(0.007)	(0.014)	(0.022)	(0.021)	(0.023)	(0.004)	(0.012)
Observations	83,169	83,169	83,169	83,169	83,169	83,169	83,169
Adjusted R^2	0.380	0.106	0.313	0.196	0.123	0.075	0.180
$\beta^M = \beta^W$ [p-value]	[0.684]	[0.000]	[0.000]	[0.849]	[0.379]	[0.368]	[0.819]
Shares	0.04	0.11	0.34	0.26	0.14	0.01	0.10

Note: This table shows the effect of the exposure to English instruction on economic activities. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in some economic activity. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Exposure to English instruction and economic industries (Mexican census data)

	(1) Agri-culture	(2) Con-struction	(3) Manu-facture	(4) Com-merce	(5) Profes-sionals	(6) Govern-ment	(7) Hospi-tality
Panel A: Full sample							
Hrs English	-0.027***	-0.028**	-0.008	0.037***	-0.014	-0.004	0.044***
	(0.006)	(0.014)	(0.011)	(0.014)	(0.015)	(0.009)	(0.014)
Observations	548,862	548,862	548,862	548,862	548,862	548,862	548,862
Adjusted R^2	0.328	0.081	0.091	0.041	0.026	0.077	0.039
Panel B: Low-enrollment sample							
Hrs English	-0.045	0.044**	-0.037	0.006	0.040***	0.024*	-0.031
	(0.080)	(0.018)	(0.041)	(0.026)	(0.011)	(0.014)	(0.032)
Observations	42,951	42,951	42,951	42,951	42,951	42,951	42,951
Adjusted R^2	0.298	0.104	0.113	0.082	0.085	0.078	0.083
Panel C: Low-enrollment sample (Men)							
Hrs English	-0.070	0.060**	-0.010	0.032	0.024**	-0.012	-0.024
	(0.089)	(0.024)	(0.048)	(0.037)	(0.011)	(0.016)	(0.018)
Observations	32,704	32,704	32,704	32,704	32,704	32,704	32,704
Adjusted R^2	0.212	0.095	0.060	0.058	0.043	0.029	0.080
Panel C: Low-enrollment sample (Women)							
Hrs English	0.062	-0.024	-0.087	-0.092***	0.091	0.101	-0.052
	(0.065)	(0.025)	(0.082)	(0.034)	(0.067)	(0.065)	(0.102)
Observations	10,247	10,247	10,247	10,247	10,247	10,247	10,247
Adjusted R^2	0.269	0.017	0.286	0.087	0.126	0.137	0.098
Shares	0.10	0.12	0.20	0.23	0.15	0.7	0.13

Note: This table shows the effect of the exposure to English instruction on economic industries. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in some economic activity. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: English exposure at county, locality and school level

	(County) Formal sector	(Locality) Formal sector	(School) Formal sector	(County) ln(wage)	(Locality) ln(wage)	(School) ln(wage)
Panel A: Full sample						
Hrs English	-0.0309***	-0.0100***	-0.0115***	-0.0345*	-0.0051*	-0.0157***
	(0.0073)	(0.0017)	(0.0015)	(0.0180)	(0.0029)	(0.0022)
Observations	15,126,879	15,126,879	15,126,879	3,580,980	3,580,980	3,580,980
Adjusted R^2	0.108	0.108	0.108	0.202	0.202	0.202
Panel B: Low enrollment sample						
Hrs English	-0.0362***	-0.0120**	-0.0125	-0.0562*	-0.0048	-0.0053
	(0.0130)	(0.0054)	(0.0084)	(0.0300)	(0.0080)	(0.0118)
Observations	1,383,624	1,383,623	1,383,624	231,139	231,139	231,139
Adjusted R^2	0.126	0.126	0.126	0.236	0.236	0.236
Panel C: Low enrollment sample (Men)						
Hrs English (β^M)	-0.0516**	-0.0156*	-0.0183	-0.0429*	-0.0006	-0.0051
	(0.0244)	(0.0082)	(0.0113)	(0.0259)	(0.0107)	(0.0180)
Observations	667,236	667,236	667,236	148,034	148,034	148,034
Adjusted R^2	0.151	0.151	0.151	0.242	0.242	0.242
Panel D: Low enrollment sample (Women)						
Hrs English (β^W)	-0.0516**	-0.0156*	-0.0183	-0.0429*	-0.0006	-0.0051
	(0.0244)	(0.0082)	(0.0113)	(0.0259)	(0.0107)	(0.0180)
Observations	667,236	667,236	667,236	148,034	148,034	148,034
Adjusted R^2	0.151	0.151	0.151	0.242	0.242	0.242
$\beta^M = \beta^W$ [p-value]	[0.2842]	[0.6220]	[0.0164]	[0.8980]	[0.2023]	[0.6991]

Note: This table illustrates the effect on the estimates to different measures of English exposure depending on the level of the observation: county, locality or school level. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 years of age. All regressions include controls and school FE. Standard errors clustered at corresponding level of exposure in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Summary statistics by occupations (Population Census 2020)

Occupations	Occupation share	Education (years)	Wage (pesos)	Formal jobs (%)
Agriculture	0.09	8.81	3,391.58	0.10
Commerce	0.16	11.13	5,062.50	0.40
Other services	0.11	11.04	5,755.06	0.41
Crafts	0.11	9.48	6,026.20	0.24
Manual work	0.18	9.81	4,792.25	0.26
Skilled job	0.13	10.41	6,348.28	0.74
Pro-Tech	0.22	12.76	6,992.73	0.60

Note: These summary statistics correspond to the sample of students who were matched to their labor market outcomes. Furthermore, these statistics represent averages of all individuals in the sample (including all six cohorts) and also over the three years of labor data (2018-2020).

Table A.6: English instruction and occupational choices
(Multinomial Logit estimates using the Mexican Population Census 2020)

	(1)	(2)	(3)	(4)	(5)
	Full	Gender		Market	
	sample	Men	Women	Formal	Informal
Agriculture					
Hrs English	0.173***	0.178***	0.172***	0.215***	0.163***
	(0.029)	(0.293)	(0.039)	(0.050)	(0.028)
Commerce					
Hrs English	1.382***	1.444***	1.267***	1.649***	1.231***
	(0.057)	(0.071)	(0.069)	(0.100)	(0.055)
Services					
Hrs English	1.550***	1.597***	1.454***	1.708***	1.435***
	(0.083)	(0.108)	(0.082)	(0.147)	(0.073)
Crafts					
Hrs English	0.917***	0.952	0.830**	0.967	0.898**
	(0.040)	(0.046)	(0.068)	(0.066)	(0.045)
Skilled job					
Hrs English	0.849**	0.935	0.676***	0.815**	1.106
	(0.054)	(0.057)	(0.064)	(0.075)	(0.072)
Professional					
Hrs English	1.623***	1.624***	1.556***	1.627***	1.657***
	(0.095)	(0.097)	(0.109)	(0.108)	(0.109)
Observations	549,672	373,798	175,874	140,699	408,973

Note: This table shows the effect of the exposure to English instruction (measured at county level) on occupational choices. Relative Risk Ratios are shown. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who participate in the labor market. All regressions include controls. My occupation of reference is all jobs requiring manual work. Standard errors clustered at county level in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Exposure to English instruction and labor market outcomes by abilities (Social Security data)

	(1) Formal Sector	(2) ln(wage)	(3) ln(distance)	(4) Move State
Panel A: Full sample				
Eng×Q2	0.001	0.000	0.008	0.000
	(0.001)	(0.001)	(0.006)	(0.001)
Eng×Q3	-0.003***	-0.002	0.016***	0.002**
	(0.001)	(0.001)	(0.006)	(0.001)
Eng×Q4	-0.017***	-0.006***	0.043***	0.006***
	(0.001)	(0.001)	(0.007)	(0.001)
Observations	15,126,882	3,580,983	3,580,983	3,580,983
Adjusted R^2	0.108	0.238	0.478	0.556
Panel B: Low enrollment sample				
Eng×Q2	0.004	-0.000	0.003	-0.010
	(0.006)	(0.008)	(0.046)	(0.009)
Eng×Q3	0.001	0.036***	0.013	-0.017
	(0.006)	(0.008)	(0.036)	(0.011)
Eng×Q4	-0.016**	0.008	0.115***	0.001
	(0.007)	(0.011)	(0.041)	(0.012)
Observations	1,383,627	231,142	231,142	231,142
Adjusted R^2	0.126	0.293	0.677	0.727
Panel C: Low enrollment sample (Men)				
Eng×Q2	0.014	0.002	-0.001	-0.003
	(0.009)	(0.011)	(0.060)	(0.010)
Eng×Q3	-0.001	0.045***	0.012	-0.010
	(0.011)	(0.013)	(0.048)	(0.011)
Eng×Q4	-0.016	0.006	0.100*	-0.000
	(0.011)	(0.017)	(0.058)	(0.015)
Observations	667,238	148,036	148,036	148,036
Adjusted R^2	0.151	0.296	0.680	0.729
Panel D: Low enrollment sample (Women)				
Eng×Q2	0.000	-0.008	0.002	-0.029**
	(0.007)	(0.013)	(0.067)	(0.012)
Eng×Q3	0.006	0.018	-0.004	-0.029
	(0.006)	(0.011)	(0.088)	(0.018)
Eng×Q4	-0.010	0.008	0.130	0.002
	(0.007)	(0.016)	(0.079)	(0.019)
Observations	716,389	83,106	83,106	83,106
Adjusted R^2	0.110	0.347	0.699	0.756
State of work FE	NO	YES	YES	YES

Note: This table shows the effect of the exposure to English instruction on labor market outcomes by quartiles of abilities. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Placebo test for pre-trends using older cohorts

	(1) Language 6th	(2) Language 6th	(3) Math 6th	(4) Math 6th
Panel A: Original estimates (1997-2002 cohorts)				
Hrs English	0.0321***	-0.0023	0.0135***	-0.0241***
	(0.0033)	(0.0075)	(0.0037)	(0.0086)
Observations	3,588,166	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.396	0.448	0.402	0.463
Panel B: Placebo test (1996-1999 cohorts)				
Hrs English	0.047***	0.020*	0.036***	-0.013
	(0.004)	(0.011)	(0.004)	(0.013)
Observations	3,141,667	3,141,667	3,141,667	3,141,667
Adjusted R^2	0.418	0.462	0.419	0.475
State FE	YES	NO	YES	NO
School FE	NO	YES	NO	YES

Note: This table tests for pre-trends comparing the original estimates with placebo results using older cohorts who where not exposed to the NEPBE. All regressions include controls and school FE. The sample contains students who later in life will work in the formal sector. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Exposure to English instruction and labor market outcomes (Social Security data)

	(1) Formal Sector	(2) ln(wage)	(3) ln(distance)	(4) Move State
Panel A: Full sample				
Hrs English	-0.008***	-0.008***	-0.002	-0.001
	(0.001)	(0.001)	(0.005)	(0.001)
Observations	15,136,808	3,588,166	3,588,166	3,588,166
Adjusted R^2	0.108	0.239	0.478	0.556
Panel B: Low ability sample				
Hrs English	-0.003	-0.006**	-0.023*	-0.003
	(0.002)	(0.003)	(0.014)	(0.002)
Observations	5,454,279	1,387,118	1,387,118	1,387,118
Adjusted R^2	0.122	0.262	0.490	0.560
Panel C: Low ability sample (Men)				
Hrs English (β^M)	-0.002	-0.008**	-0.036**	-0.005*
	(0.003)	(0.004)	(0.018)	(0.003)
Observations	2,818,675	899,295	899,295	899,295
Adjusted R^2	0.129	0.259	0.500	0.567
Panel D: Low ability sample (Women)				
Hrs English (β^W)	-0.004	-0.000	0.003	0.001
	(0.003)	(0.004)	(0.023)	(0.004)
Observations	2,635,604	487,823	487,823	487,823
Adjusted R^2	0.104	0.308	0.501	0.572
$\beta^M = \beta^W$ [p-value]	[0.000]	[0.000]	[0.000]	[0.001]
State of work FE	NO	YES	YES	YES

Note: This table shows the effect of the exposure to English instruction on labor market outcomes. The sample contains all Mexican workers who belong to the cohorts 1997-2002, who are less than 22 and who are employed in the formal sector. All regressions include controls. Standard errors clustered at school level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$