BILINGUAL EDUCATION:
EXPERIENCE FROM MADRID

Miguel Ruiz

Master Thesis CEMFI No. 1701

May 2017

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551. Fax (34) 914 291 056
Internet: www.cemfi.es

I am greatly indebted to Caterina Calsamiglia for her guidance and to Manuel Arellano for his time and selfless dedication. I am very grateful to Joyce Dvorak for her assistance with this research. I thank Luis Pires, Ismael Sanz, the Consejería de Educación de la Comunidad de Madrid, Brindusa Anghel, Antonio Cabrales and Jesús Carro for letting me access the database. I also want to thank CEMFI's faculty for useful comments and suggestions. I acknowledge the financial support received from CEMFI during this project and the completion of the 2014-2016 Master in Economics and Finance. All errors are my own.

Master Thesis CEMFI No. 1701
May 2017

BILINGUAL EDUCATION: EXPERIENCE FROM MADRID

Abstract

Bilingual education programs promote students' language proficiency and communicative competence in a language other than their own. Nowadays, bilingual programs are present worldwide responding to an increasing demand partially driven by the potential personal and economic benefits from being proficient in a foreign language. However, bilingual education increases the difficulty of learning academic content due to it being taught in a non-native language. To measure the importance of this effect, I utilize standardized test data and the Spanish-English bilingual program of Madrid. The findings show a small but significant negative impact of the program on the performance of students in English-taught content. The negative effect is stronger around the median of the student's ability distribution.

Miguel Ruiz
CEMFI
mruiz@cemfi.edu.es

1 Introduction

Bilingual education programs exist worldwide in a variety of countries including Argentina, India, the Netherlands, the United States of America and Spain. They offer academic content instructed in different languages to promote the lifelong improvement of communicative skills in two different languages. These programs respond to an increasing demand for proficiency in a foreign language enhanced by an integrated global market. The widespread knowledge of languages is an important determinant of foreign trade with English playing an important role as shown by Fidrmuc and Fidrmuc (2009). Furthermore, there is a positive link between English proficiency and success in the labor market (Gonzalez, 2000; Trejo, 2003; Bleakley and Chin, 2004). In the case of Spain, the returns to the English language have been found to be very large through Mincerian regressions; Ginsburgh and Prieto-Rodriguez (2011) find point estimates of a 39% increase in wages.

The typical objectives of bilingual programs include language, academic and affective dimensions. High levels of proficiency in their first and a second language, performance at or above grade level in academic areas in both languages and positive cross-cultural attitudes and behavior[1](Christian, 1994). Additionally, a number of studies have reported that bilingual children exhibit a greater sensitivity to linguistic meanings and may be more flexible in their thinking than monolingual children (Diaz, 1985; Cromdal, 1999). By becoming competent in two language systems, the bilingual student has had to decipher much more language input than the monolingual student. Therefore, the bilingual students has had more practice in analyzing meanings than the monolingual student (Cummins, 2000).

The term "Bilingual Education" (BE) is inherently ambiguous as it encompasses both submersion and immersion programs. The term "submersion program" refers to a situation where language minority children are taught in the language of the general population. This type of instruction aims to develop the command of a language that might be foreign to the minority but is the dominant language outside the school. The impact of bilingual programs of this type has been widely studied in the USA. For example, Jepsen (2010) studies the impact of this type of bilingual program on Spanish-speaking, English learning students in California[2]. Also, a meta-analysis on the effectiveness of submersion bilingual programs can be found in Greene (1998). Alternatively, BE also refers to a situation in which a language which is not the dominant language for the larger society is the medium of instruction. This situation is referred to as "immersion". The impacts of bilingual immersion programs in Europe have been studied by Admiraal, Westhoff and de Bot (2006) in the Netherlands as well as by Anghel, Cabrales and Carro (2015) and Gerena and Verdugo (2014) in Madrid.

Bilingual education increases the difficulty of learning academic content due to it being taught in a foreign language. Thus, the potential negative consequences such additional hardship might have on the development of student's language skills and academic knowledge are a concern.

[1]The positive cross-cultural externalities coming from a bilingual program are to be understood in a context where there is a language minority; in this case Spanish speakers in California. Thus, a bilingual program in this setting also aims to foster cross-cultural interactions between native Spanish and English students under a bilingual Spanish-English framework.

[2]An English learning Student (ELs) is a student who is not proficient in English and speaks a language other than English at home.

In regards to linguistic competencies, the evaluation of bilingual programs for both majority and minority language students shows no long-term academic delay in the majority language for those students exposed to partial minority language instruction (Appel and Muysken, 2006). This result can be interpreted as evidence that the first and second language academic skills are interdependent and manifest a common underlying proficiency (Cummins, 2000). These cross-lingual relationships are stronger between languages with more commonalities as shown by Genesee (1987). Regarding academic content, I am interested in the impact of BE on the knowledge acquired through English instruction for native Spanish speaking students within the Autonomous Community of Madrid. This paper, looks into the Spanish-English bilingual program in Madrid, where the two main languages of instruction are not as closely related as the languages in other bilingual programs in Spain such as the Catalan-Spanish or the Galician-Spanish bilingual education programs in their respective autonomous communities. In 2004, the regional Ministry of Education in Madrid launched the *Madrid Comunidad Bilingüe* program (MCB). This program is a Spanish-English bilingual immersion program where students receive at least 45% of the academic instruction in English.

The aim of this paper is to contribute to the empirical literature on bilingual education programs by studying the impact of BE on content learned in a non-native language. To do so, I look into the bilingual program of Madrid together with a standardized test students in Madrid sit at the end of primary education. My contribution is twofold. On the one hand, I estimate the average treatment effects of the bilingual program on the performance of students on the academic content instructed in English. The former research on this policy conducted by Anghel, Cabrales and Carro (2015) used a standard Diff-in-Diff specification. In this paper, a Diff-in-Diff specification with school fixed effects is used. The model presented controls for unobservable time-invariant characteristics at the school level including student school-average unobservable characteristics as long as there is not much time-variation in these average characteristics. The analysis is conducted using newly available data for the years 2013 and 2015. There is statistical evidence that the increased difficulty of learning in a foreign language has a negative and significant impact on content learning. On the other hand, the distributional effects of the same policy are studied modifying the 2 step method proposed by Chetverikov, Larsen and Palmer (2016) exploiting the extra time dimension available in the data. Their method is of practical significance when the researcher has data on a group-level endogenous treatment and has microdata on the outcome of interest within each group. Specifically, I follow the same first step as the authors but use a different specification for the second step instead of OLS/IV to address the expected endogeneity of the program. The results show that the impact of the policy is concentrated around the median with smaller effects at the tails of the distribution.

This paper does not evaluate the bilingual program of Madrid as a whole but solely focus on the impact of the increased difficulty of learning content in a non-native language. Future research, with newly available data, will also be able to provide with an estimation of the expected positive impact of the policy in students' English proficiency which is the main purpose of this policy and that I do not investigate. By no means does this paper conclude that the negative impact found in the content taught in English is an indication that the bilingual program in Madrid is not successful.

The paper is organized as follows. In Section 2 the institutional background and program information relevant to this study are provide. In Section 3 the standardized test data utilized to conduct this paper's analysis is described. In Section 4 the empirical strategy followed to analyze both average and distributional effects is presented. Estimation results are provided in Section 5. Conclusions and limitations of the analysis are presented in Section 6 together with policy recommendations. Section 7 displays the different robustness checks for the average and distributional estimation results.

2 The Bilingual Program in Madrid

2.1 Purpose and Nature of the Program

The *Madrid Comunidad Bilingüe* program is a Spanish-English bilingual program which promotes the acquisition and lifelong improvement of communicative skills in English responding to a primary objective of the European Union. State bilingual primary schools, in addition to the English language classes, teach at least three other subjects in English, excluding Mathematics and Spanish Language & Literature.

The program can be categorized as a Content and Language Integrated Learning (CLIL) program. Under this methodology, a secondary language is used as the medium of instruction for some academic content. Thus, it serves the dual purpose of teaching the same specific curricular content as non-bilingual primary schools in Madrid while also improving the overall English proficiency. This kind of approach has been identified as very important by the European Commission as: "It can provide effective opportunities for pupils to use their new language skills now, rather than learn them now for use later. It opens doors on languages for a broader range of learners, nurturing self-confidence in young learners and those who have not responded well to formal language instruction in general education. It provides exposure to the language without requiring extra time in the curriculum, which can be of particular interest in vocational settings." (European Union Council, 2008).

The MCB program began in 2004 with 26 state bilingual schools and has grown to 352 public primary schools and 181 charter schools. In 2010, 32 state bilingual secondary schools joined the program and the number has increased to 110 in 2016.

Table 1: Number of Primary Schools and Students under the Bilingual Program

Year	*2005*	*2007*	*2008*	*2010*	*2015*
Primary Schools	26	122	147	206	336
Students	1481	10949	18439	37765	88000

2.2 Selection of New Bilingual Schools

Since 2004, the Regional Ministry of Education has published a yearly order containing an official call for the selection of new schools to join the bilingual program. This order establishes the requirements for schools to be able to guarantee that they are able to implement a bilingual program. When the program launched in 2004, there were three criteria used to evaluate the applicants to the program:

1) *Acceptance from the educational community*: Given by the application to the program by both the school teachers and the school board. The school board is a decision making body formed by the school's principal, teachers and elected parents.

2) *Feasibility of application*: The school's former experience in relevant teaching practices through small scale pilot programs, teacher specialization in English and school resources, number of students and classes.

3) *Geographical equality*: The first schools under the MCB program were selected so as to have a balanced distribution of bilingual schools across the entire Madrid Community.

In 2005, the third requirement was substituted by

3') *Teacher's Proficiency in English*: Teachers are required to show credentials ensuring their command of English to be sufficient to properly instruct in the bilingual program[3].

2.3 Elements of the Bilingual Program

1) *Teaching staff*

   - The principal and bilingual program coordinator: Together with the school management team the principal and bilingual program coordinator collaborate in the organization of teaching assistants and the external evaluation.
   - Primary and secondary school teachers: In recognition for their efforts to teach in a foreign language and the obtained Linguistic Certificate, they receive a financial bonus in accordance with the number of hours taught in English.

2) *Teacher training*: The Regional Ministry of Education provides training to teachers, principals and school management teams. These training programs take place abroad in English-speaking countries with the goal of improving linguistic as well as managerial skills.

3) *Certificate to teach in a foreign language:* Teachers in the bilingual program must first obtain the certification required for bilingual teaching positions. There are two ways to become certified:

   a) Submit certain university degrees or official language certifications deemed to be equivalent to a CEFRL[4] level C1 or above.

   b) Pass an exam which recognizes – exclusively for teaching positions in the Autonomous Community of Madrid – a C1 level of linguistic proficiency.

[3] A more detailed description of the required English credential can be found in Section 2.3
[4] Common European Framework of Reference for Languages

4) *Language assistants:* Language assistants are university graduates from English speaking countries who support the work of teachers in the classroom. The selection of these individuals is done jointly by the Regional Ministry of Education, the (National) Ministry of Education, the Fulbright Commission and the British Council. Language Assistants are accepted for one academic school year and they can apply to renew their contract for a second year. They collaborate 16 hours per week at their assigned schools. In return, they receive a stipend from the Regional Ministry of Education which amounts to 1,000 euros a month during the academic year. During the 2014-2015 academic year there were 1,098 English speaking language assistants assigned to state primary schools and 428 to state secondary schools[5].

5) *External evaluation:* The students in the bilingual program have their English skills assessed in second, fourth and sixth grade of primary school and during the second and fourth years of state secondary school. These external evaluations are administered by the British Council.

3 Data

Every year, the Community of Madrid administers the standardized test *Conocimientos y Destrezas Indispensables* (CDI). This exam takes place annually at two different ages. The primary school CDI takes place in 6th grade (sexto de primaria) and the secondary school CDI takes place in 9th grade (tercero de ESO). Both exams are conducted in Spanish, regardless of whether students are enrolled in a bilingual program. The exam is undertaken by every student in Madrid, so it collects information from public, charter and private schools. Similar to PISA, the primary school CDI has no academic repercussions for the student beyond the information they provide the school, parents and students about their relative performance[6].

The primary school CDI exam is divided into two parts. The first part lasts an hour and evaluates the students' skills in Reading, Grammar, and General Knowledge[7]. The contents of the General Knowledge part correspond to the contents covered in the classes of Social and Natural Science[8]. The second part lasts 45 minutes and evaluates Mathematical knowledge[9]. Besides the scores for the different parts of the CDI exam, the data set also provides three basic controls available for each student: gender, age and nationality.

3.1 Data Structure

The data utilized contains student-level data grouped by schools on outcomes of interest plus controls and school-level treatment data. The dependent variables of interest are the scores obtained in Mathematics, Language and Science. Moreover, a number of indices are constructed

[5]In addition to English-speaking language assistants, there are also 39 German and French teaching assistants in the Autonomous Community of Madrid.

[6]The secondary school CDI exam, which used to take place in 9th grade, has now been removed. The grade obtained in the secondary education CDI test was more than just informative as it would be a part of the eligibility criterion for the Excellence Program in Madrid

[7]Hereafter, General Knowledge will be referred to as Science.

[8]In some schools, both subjects are merged into a single subject called "Conocimiento del Medio"

[9]The Primary School CDI carried out in 2009, 2010 and 2011 also answered to a survey which provided with rich socio-economic information about the students. For unknown reasons, the survey was suspended until 2015

using question specific data available for the years 2013 and 2015. The primary school CDI data used comes from the years 2013 and 2015 although CDI results from 2009 to 2011 are used in several robustness checks. Both dependent and control variables including gender, nationality and age are available at the student-level.

4 Empirical Strategy

In this section, a brief introduction to the methodology used in quasi-experimental research is provided. The first subsection may be unnecessary to readers familiar with econometric theory but will hopefully provide with the basic intuitions to those not so familiar with this type of estimation tools. First, the (unfeasible) ideal experiment which would capture the causal relationship of interest is presented together with the econometric specification which could be used to estimate the causal effect of interest. Then, given the reality the researcher confronts, the different models used to estimate both average and distributional effects of the bilingual policy in Madrid are explained.

4.1 The Ideal Experiment

The purpose of this paper is to estimate the impact (causal effect) of the program *Madrid Comunidad Bilingüe* (MCB) on the performance of students in content learned in a foreign language. To capture the causal effect of the bilingual program on student performance, we could rely of the following ideal experiment. Once schools apply to join the program, we could randomly select treatment and control schools conditional on observable characteristics. All the schools participating in the bilingual program would have exactly the same subjects and hours taught in English and external evaluations were to be conducted in both treatment and control schools in a variety subjects not just English. There would be a record of student transfers to properly control for migration between schools and a number of socioeconomic covariates including parents' education level and occupation.

Consider Y_{1i} and Y_{0i}, the potential outcomes of individual i. In other words, Y_{1i} is the grade of a student had she gone to a bilingual school, irrespective of whether she actually went to a bilingual school, while Y_{0i} is the individual's grade if she had not attended a bilingual school. We would like to know the difference between Y_{1i} and Y_{0i} which could be identified as the causal effect of attending a bilingual school for individual i. However, it is impossible to observe both potential outcomes for individual i. Therefore, it is necessary to rely on comparisons between those students who attended a bilingual school and those who did not. The observed outcome, Y_i, can be written in terms of potential outcomes.

$$Y_i = \begin{cases} Y_{1i}, & \text{if } MCB_i = 1 \\ Y_{0i}, & \text{if } MCB_i = 0 \end{cases} \qquad \text{or} \qquad Y_i = Y_{0i} + (Y_{1i} - Y_{0i})MCB_i \tag{E1}$$

Where MCB_i is a dummy equal to 1 if student i attended a bilingual school.

Conditional on observable characteristics, the difference between Y_{1i} and Y_{0i}, which could be identified as the causal effect of attending a bilingual school for individual i, is given by α.

$$\alpha = E[Y_{1i} - Y_{0i}|X_i] = E[Y_{1i} - Y_{0i}|MCB_i = 1, X_i] = E[Y_i|MCB_i = 1, X_i] - E[Y_i|MCB_i = 0, X_i]$$

Where E is the expectation function. In other words, $E[Y_{1i} - Y_{0i}|X_i]$ is the expectation of the causal effect given the observable characteristics X_i.

If the effect of the policy is the same for everybody, we say there are constant treatment effects. The following linear regression can be used to estimate the causal effect of the policy on students' performance given by equation (E1).

$$Y_i = \beta_0 + \alpha\ MCB_i + \delta' X_i + \eta_i \tag{1}$$

Where Y_i is the exam grade for student i, MCB_i is a dummy if the school belongs to the treatment/bilingual group and X_i are observable student characteristics. α is the parameter of interest which captures the causal effect of the bilingual policy on students' grades. δ' is a vector parameter capturing the effect of the different student-individual covariates.

4.2 The Identification Strategy

The linear regression proposed for the ideal experiment would lead to biased estimates as the implementation of the bilingual program was done as described in Section 2. Recall that the school selection was not random and that we do not have data on student socio-economic background or transfers. Therefore, if we estimate the causal effect using the method proposed in Section 4.1 we would be suffering from selection bias at both school and student level. In other words, if better performing schools are more likely to implement the bilingual program, the estimated effect will be biased upwards as it will be contaminated by the fact that bilingual schools perform better to begin with. By the same token, if students/parents with higher income levels, associated with better academic performance, are more likely to join bilingual schools, our estimation will biased upwards.

In order to extract the causal effect of the MCB program given the reality we confront, we need to rely on econometric methods to mitigate these sources of bias. The key to the causal effect estimation in Section 4.1 was to control for confounding factors such as students background, transfers or school-specific characteristics. Nevertheless, even if we had many controls, we still expect important confounders to be unobserved. But, we also expect some of these unobserved confounders to be fixed so we will rely on methods which utilize the cohort dimension of the available data. The underlying identification assumption in these models is that the counterfactual trend for both treatment and control groups is the same. Conditional on the observed characteristics, the trends in the performance of both treatment and control schools must be the same.

4.3 Statistical Inference

4.3.1 Estimating Average Effects

To estimate the average treatment effects, I use yearly observations by pairs grouped by schools to have a before-&-after framework with student level outcomes of interest and covariates. The treatment group is constituted by the bilingual schools that were bilingual in 2015 but not in 2013. The control group is constituted by the non-bilingual schools[10]. I estimate the following equation for $t = \{2013, 2015\}$

$$Y_{ist} = \eta_s + \theta_t + \gamma_{DID}(D_t \times MCB_s) + \delta' X_{ist} + u_{ist} \tag{2}$$

Where Y_{ist} is the CDI exam grade available for Mathematics, Language and Science or the grade in the index constructed using individual questions. η_s is the school fixed effect, θ_t is the year fixed effect, MCB_s is a treatment school dummy and D_t is a year dummy. X_{ist} are observable student characteristics including age and dummies for gender and Spanish nationality. γ_{DID} is the parameter of interest capturing the effect of the bilingual policy on students' grades. The estimated $\hat{\gamma}_{DID}$ will measure the impact of the program on the students' performance on the dependent variable.

4.3.2 Indices built with individual questions

The CDI data for the years 2013 and 2015 contain question specific observations at the student level which can be used to construct dependent variables Y_{ist} to assess whether the negative impact found in Science responds to “knowledge” or “knowledge in Spanish”. Recall that the CDI test is conducted in Spanish regardless of whether the students attend a bilingual school. The specific questions do not have a numerical value but have three possible scores: ”Good”, ”OK” or ”Wrong”. The numerical values 1, 0.5 and 0 are assigned respectively. The constructed indices are displayed in Section 5.2. These variables are: “Common”: self-elaborated indicator using only those questions belonging to a field present in both exams[11]. “No Vocabulary”: Average science grade excluding the vocabulary intense questions. “Geography”: Average of Geography questions. “Vocabulary”: Average grade of vocabulary intense questions.

4.3.3 Estimating Distributional Effects

We are also interested in the effect of the policy on the within-school distribution of student-level outcomes. We would like to know which part of the ability distribution of students is driving the estimated average effect on Science[12].

[10] When a school joins the bilingual program, it takes 5 years for their first bilingual cohort to be exposed to the CDI standardized test. To avoid confusion, I will generally refer to bilingual schools as those facing the CDI with a bilingual cohort. For example, the bilingual schools facing the 2015 CDI test with a bilingual cohort are schools that joined the bilingual program in 2010.

[11] The 2013 exam has a question regarding musical instruments and Health which do not have a content match in 2015. The 2015 exam has 2 questions on History which do not have a counterpart in 2013.

[12] Given the available student covariates, the distribution of ability could be understood as the distribution of students' suitability to study including their ability, situation at home, parents educational level and study environment among other unobservable factors.

To estimate the distributional effects of the policy on students' performance in Natural and Social Science I will focus only in the years 2013-2015 utilizing the same treatment and control groups as in Model (2). The treatment group will be constituted by the bilingual schools that were bilingual in the year 2015 but not in 2013. The control group will be constituted by the non-bilingual schools. For this estimation, I will use a modified version of the two-step method proposed by Chetverikov, Larsen and Palmer (2016) to exploit the extra time dimension available in the data.

Step 1: Quantile regressions within each school are performed to estimate the effects of micro-level covariates on individual student outcomes.

$$\mathrm{Q}^s_{y_{it}|x_{it}}(\tau) = \alpha^s_t(\tau) + \beta'^s_t(\tau)\, x_{it} + \epsilon^s_{it}(\tau), \quad \tau \in \mathcal{T} \text{ and } s \in \mathcal{S} \tag{3}$$

Where $\mathrm{Q}^s_{y_{it}|x_{it}}(\tau)$ is the τth conditional quantile of y_{it} in school s, y^s_{it} is the outcome of interest at the student level in school s which is given by the CDI grade in Mathematics, Language or Science. x_{it} are student level covariates including age and dummies for gender and Spanish nationality. $\mathcal{T}$ is the set of quantile indices of interest and $\mathcal{S}$ is the set of schools.

$\hat{\alpha}^s(\tau)$ is estimated from equation (3) . Where $\hat{\alpha}^s(\tau)$ can be understood as the school level quantile net of the effect of student observable characteristics.

$$\hat{\alpha}^s(\tau) = \frac{1}{N_s}\sum_i \left(\hat{\mathrm{Q}}^s_{y_i|x_i}(\tau) - \hat{\beta}^s(\tau)\, x_i\right)$$

Where N_s is the number of students per school.

Step 2: $\hat{\alpha}^s(\tau)$ is regressed on school level covariates, including a treatment dummy, using a suitable specification to capture the effect of the bilingual program. By doing so, we capture the effect of the policy on student grades net of student characteristics for each quantile. Recall that the school acceptance into the program was not done at random and thus I expect the bilingual schools to self-select into the program. In other words, the treatment dummy included in the vector of school characteristics X_s is expected to be endogenous. A Difference-in-Difference specification with school fixed effects alike the one performed for the average treatment effects is used[13]. The second step of the methodology proposed by Chetverikov, Larsen and Palmer (2016) is modified in order to exploit the extra time dimension in the data available. Consider the following specification:

$$\alpha_{st}(\tau) = \eta_s(\tau) + \theta_t(\tau) + \psi(\tau)(D_t \times MCB_s) + \delta'(\tau)X_{st} + u_{st}(\tau) \tag{4}$$

Where η_s is the school fixed effect, $\theta_t(\tau)$ is the year fixed effect, MCB_s is a treatment school dummy, D_t is a year dummy and X_{st} are observable school characteristics. $\psi(\tau)$ is the parameter of interest measuring the impact of the bilingual policy on students' grades for each quantile τ.

[13]Chetverikov, Larsen and Palmer (2016) regress the estimated $\alpha_s(\tau)$ on group-level characteristics using OLS assuming $X_s \perp u_{st}$. However, the authors also indicate that if X_s is believed to be endogenous, an IV approach can be followed using Hausman and Taylor (1981) proposed methodology.

5 Results

In this section, the results for the different specifications introduced in Section 4 are presented. Firstly, the econometric specification is shown again followed by the descriptive statistics for the treatment and control groups. Finally, the estimation output for the different models is provided and the estimated parameters of interest explained.

5.1 Average Effect on Mathematics, Language and Science

The model used to estimate the average effect of the bilingual program in students' performance was given by equation (2) for $t = \{2011, 2013\}$ and $t = \{2013, 2015\}$:

$$Y_{ist} = \eta_s + \theta_t + \gamma_{DID}(D_t \times MCB_s) + \delta' X_{ist} + u_{ist} \tag{2}$$

Where γ_{DID} is the parameter of interest capturing the effect of the bilingual policy on students' grades. Tables (12) and (2) display the summary statistics for the treatment and control schools for the years 2013 and 2015. The groups are very similar in terms of outcome variables and covariates.

Table 2: Descriptive Statistics - 2013 & 2015

Mean / (SD)	2013		2015	
	Treatment	Control	Treatment	Control
Mathematics	7.097	6.761	7.179	6.810
	(2.260)	(2.399)	(2.521)	(2.663)
Language	8.032	7.804	7.707	7.531
	(1.842)	(1.983)	(1.741)	(1.941)
Science	8.273	8.033	6.141	6.135
	(1.658)	(1.811)	(2.316)	(2.403)
Female	0.475	0.494	0.482	0.491
	(0.499)	(0.500)	(0.500)	(0.500)
Spanish	0.931	0.893	0.955	0.919
	(0.254)	(0.309)	(0.208)	(0.273)
Age	12.099	12.145	12.081	12.133
	(0.309)	(0.365)	(0.294)	(0.355)
Observations	6,128	46,588	6,618	48,786

Table 3: Equation (2) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	**Mathematics**	**Language**	**Science**
Diff-in-Diff	0.047	-0.058	-0.233**
	(0.077)	(0.062)	(0.094)
Female	-0.298***	0.181***	-0.323***
	(0.015)	(0.012)	(0.013)
Spanish	0.096***	0.323***	0.356***
	(0.033)	(0.030)	(0.030)
Age	-2.073***	-1.703***	-1.679***
	(0.028)	(0.025)	(0.024)
Constant	32.025***	28.126***	28.285***
	(0.343)	(0.300)	(0.296)
Observations	105,684	105,668	105,668
Adjusted R-squared	0.249	0.241	0.383

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

The estimates obtained for equation (2) for $t = \{2013,\ 2015\}$ can be found in Table (3). There is evidence that girls perform relatively better in Language but worse in Mathematics and Science than boys. Spanish students perform relatively better than the rest of students in the three subjects and older students do worse in all three subjects. The estimation of the parameter of interest γ_{DID} capturing the impact of the bilingual program on Science is negative and significantly different from zero. The bilingual school program has no significant effect on either of the two subjects instructed in Spanish. The results show a similar magnitude to the estimates obtained by Anghel, Cabrales and Carro (2015) in their study evaluating the bilingual program in Madrid for the first two bilingual cohorts in the years 2009 and 2010.

5.2 Indices Built from Individual Questions

The model used to estimate the average effect of the bilingual program in students' performance on indices built with specific questions is given by equation (2) for $t = \{2013, 2015\}$:

$$Y_{ist} = \eta_s + \theta_t + \gamma_{DID}(D_t \times MCB_s) + \delta' X_{ist} + u_{ist} \tag{2}$$

Where γ_{DID} is the parameter of interest capturing the effect of the bilingual policy on students' grades. Table 4 displays the summary statistic for the treatment and control groups. Table 5 displays the estimation results of equation (2). The estimated effect is negative and significant if we consider the index constructed using only questions with a content match in both years. The results are also negative and significant if we use the index excluding vocabulary intense questions or Geography questions. The findings show no significant effect on vocabulary intensive questions.

Table 4: Question Specific Refinement Summary Statistics for $t = \{2013, 2015\}$

Mean / (SD)	2013 Treatment	2013 Control	2015 Treatment	2015 Control
Science	8.032	7.804	6.141	6.135
	(1.842)	(1.983)	(2.316)	(2.403)
Common	6.990	6.649	6.594	6.527
	(2.539)	(2.661)	(2.391)	(2.506)
No Vocabulary	7.156	6.811	6.072	6.091
	(2.660)	(2.786)	(2.779)	(2.844)
Geography	8.892	8.677	5.037	5.199
	(1.941)	(2.134)	(3.156)	(3.213)
Vocabulary	6.491	6.162	8.157	7.835
	(3.802)	(3.904)	(2.932)	(3.195)
Female	0.475	0.494	0.482	0.491
	(0.499)	(0.500)	(0.500)	(0.500)
Spanish	0.931	0.893	0.955	0.919
	(0.254)	(0.309)	(0.208)	(0.273)
Age	12.099	12.145	12.081	12.133
	(0.309)	(0.365)	(0.294)	(0.355)
Observations	6,128	46,585	6,618	48,785

Table 5: Equation (2) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Science	Common	No Vocabulary	Geography	Vocabulary
Diff-in-Diff	-0.239**	-0.276***	-0.367***	-0.375**	-0.001
	(0.102)	(0.096)	(0.110)	(0.149)	(0.150)
Female	-0.086***	-0.219***	-0.221***	-0.594***	-0.213***
	(0.014)	(0.016)	(0.018)	(0.018)	(0.022)
Spanish	0.306***	0.458***	0.291***	0.277***	0.961***
	(0.031)	(0.034)	(0.036)	(0.036)	(0.049)
Age	-1.817***	-2.009***	-2.108***	-1.634***	-1.712***
	(0.025)	(0.028)	(0.030)	(0.030)	(0.037)
Constant	29.658***	30.774***	32.288***	28.583***	26.233***
	(0.311)	(0.340)	(0.372)	(0.363)	(0.462)
Observations	105,668	105,668	105,668	105,668	105,668
Adjusted R-squared	0.338	0.243	0.245	0.441	0.176

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

5.3 Distributional Effect on Science

The model used for the estimation of the distributional effects of the bilingual program on the students' performance on Science is given by equation (4):

$$\alpha_{st}(\tau) = \eta_s(\tau) + \theta_t(\tau) + \psi(\tau)(D_t \times MCB_s) + \delta'(\tau)X_{st} + u_{st}(\tau) \tag{4}$$

The descriptive statistics for the treatment and control groups can be found in Table (2). Table (6) shows the estimation results for the years 2013-2015. The estimation of the parameter of interest $\psi(\tau)$ capturing the impact of the bilingual program on Science is found to be stronger around the median. The estimated effects are negative and significantly for the 0.3, 0.4, 0.5 and 0.6 quantiles. An extended version of this table is provided in Table 24. Moreover, a graphical representation of the results can be found in Figure (1). The magnitude of the estimates can no longer be read in the units of the grade as it was the case for the average treatment effects. The average treatment effect calculated in Section 5 was -0.233 and the average treatment effect calculated through the 2 step method is -6.082. Using a back-of-the-envelope calculation, we could approximate this magnitude in terms of the exam grade to be -0.331 points less on the 0-10 exam grade scale.

The estimation results for $\delta'(\tau)$ indicate peer effects at the school level coming from the proportion of Spanish students in class. The peer effect results are similar to the ones obtained by Silaghi (2011) using data also from the CDI standardized test[14]. The proportion of Spanish students in the school is found to have a positive and significant effect on student performance.

Table 6: Equation (4) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Science				
	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Diff-in-Diff	-2.987	-7.606*	-8.650**	-4.232	-3.742
	(4.777)	(4.127)	(3.555)	(3.845)	(4.139)
Female	1.725	-7.483	-4.106	-9.599	9.337
	(10.600)	(9.091)	(8.522)	(9.158)	(10.821)
Spanish	13.962	29.566**	28.287***	27.234**	24.723**
	(14.576)	(12.173)	(10.356)	(12.684)	(12.505)
Age	54.948***	25.480**	16.396	16.052	-6.657
	(14.708)	(12.815)	(12.407)	(13.208)	(15.183)
Constant	-657.986***	-304.749*	-195.356	-188.293	80.324
	(181.683)	(158.725)	(153.888)	(163.457)	(188.381)
Observations	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.147	0.080	0.093	0.062	0.090

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

6 Conclusions

In this section, the conclusions drawn from the results displayed in Section 5 are presented. First, the interpretation of the estimates are provided. Second, comments on the limitations this analysis faces are explained. Finally, policy recommendations which would be helpful to facilitate the evaluation of this educational policy are provided.

[14] Silaghi (2011) finds evidence of a negative and significant impact of the fraction of immigrants in schools on the academic achievement of native pupils.

6.1 Main Results

The estimated average effects show that there is statistical evidence that the bilingual program has a negative and significant impact on the performance of students in the subjects instructed in English. The same models show no significant effect on the two subjects instructed in Spanish. These results are robust to falsification tests, different time periods, different treatment and control groups and different specifications as shown in Section 7.1.

The indices built using question specific data display an effect not significantly different from zero for vocabulary-intense questions while the effect is negative and significant for the rest of indices, including Geography. Thus, there is statistical evidence that the negative and significant effect found in Science does not respond to students lacking content specific vocabulary in Spanish but rather a small but significant lack of content knowledge.

The negative effect found in Science is found to be driven by the distribution of students around the median with a lower effect at the tails of the distribution. The distributional results are robust to alternative specifications as shown in Section 7.2. Additionally, Section 7.3 includes a non-linear model specification.

6.2 Limitations

6.2.1 Student Selection Bias

In Section 4.2 the identification strategy followed to estimate the causal effect of interest was described. Exploiting the cohort dimension of the data set for different years, school fixed effects are included to capture unobservable time-invariant school-level characteristics. This identification controls for the self-selection of schools into the program but does not address student/parent self-selection fully. It would address this selection bias as long as there is not much variation in average school-level socioeconomic characteristics.

Regarding student/parent selection bias, there are at least two potential channels through which student self-selection might be biasing the estimation. On the one hand, parents with better socio-economic background may have more flexibility in terms of mobility or devote more time to carefully choose their children school. By strategically choosing their residence location across the city, highly educated parents could systematically be changing residence prior to pre-school years in order for their children to have better chances to attend a specific type of public school. The specification with school fixed effects controls for this selection bias as long as the average parent's socio-economic characteristics in each school do not vary much between 2015 and 2013 in response to the bilingual program implementation. On the other hand, there could be student transfers throughout the 6 years of bilingual primary education induced by the program itself biasing our results. If students in bilingual schools with low academic performance systematically move to non-bilingual schools to avoid the increased difficulty of learning in a foreign language then our estimates could be biased upwards and the estimated effect could in fact be more negative. In Section 7.1.2, a robustness check excluding repeaters is conducted in order to account for this effect. By doing so, I am only able to control for those students who, driven by very low academic performance, may transfer to a non-bilingual school when they have to retake an academic year. However, I do not control for low performers who are not doing bad enough to retake an academic year but who might decide to transfer to a non-bilingual school to avoid the increased difficulty.

6.2.2 What is the estimated effect on Science measuring?

The CDI standardized test is undertaken in Spanish regardless of the school the students attend. For this reason, the negative effect found in Science may not reflect the cognitive ability of students but the fact that the students do not know the Spanish translation of the concepts they are instructed in English. Thus, what does the estimated effect measure? *"Knowledge"* or *"Knowledge in Spanish"*? To answer this question, several indices were constructed using question specific data in Section 4.3.2. Although the results show some evidence that vocabulary in a foreign language does not seem to be driving the results, the unsophisticated indices used to measure these subtle aspects are far from being satisfactory.

6.2.3 Teaching Practices

It can be argued that results from standardized test might capture different types of student cognitive ability. One concern could be that the change in teaching practices induced by bilingual programs could lower standardized test performance while having a non-significant or even positive impact on student cognitive ability. Consider the following two teaching practices: traditional and modern teaching practices. Traditional teaching practices rely on lectures and repetitive practice for student learning. However, since the 20th century there have been several movements towards a more student-centered educational approach where teamwork and discussion among students are given a lot of importance. Modern teaching practices encompass the latter. Traditional teaching practices have been associated with better performance in standardized tests (Lavy, 2015; Schwerdt and Wuppermann, 2011). Alternatively, modern teaching practices have been associated with smaller or even negative impact on standardized test performance (Lavy (2015), Murnane and Phillips (1981) and Goldhaber and Brewer (1997)). Beyond the performance in standardized tests, the impact of teaching practices on student's congitive ability has been studied in Bietenbeck (2014).

The bilingual program in Madrid provides native English speakers as teaching assistants for those classes instructed in English. The introduction of teaching assistants into the classroom could potentially be inducing a shift from traditional to modern teaching practices. Controlling for such change can be crucial in order to assess the real impact of the bilingual program on the student knowledge in English-instructed subjects.

6.2.4 External Validity and Teacher Quality

The main disadvantage for the treatment effects approach is that the estimated parameters are not deep-parameters (they are reduced-forms) and as a result they are not policy-invariant (Lucas (1976); Heckman and Vytlacil (2005)). More importantly, the results obtained might be an indication of the potential impact of the policy but the subsequent extension of the program might bring unwanted and unexpected results not predicted by the quasi-experimental research. To illustrate this phenomenon, one good example is the potential student achievement gains from smaller class size. The benefits of the class-size reduction programs (CSR) have been well documented over the past two decades (Angrist and Lavy (1997); Krueger and Whitmore (2001); Krueger (2003)). As a result of the positive and promising estimates obtained by a number of experimental and quasi-experimental studies, national-wide CSR programs were adopted throughout the United States in 2005. Surprisingly, the follow-up studies of the effect of these policies once extended show mixed results which vary from negative to short-comings of what the former small scale studies had suggested. One possible explanation for these discouraging results was provided by Dieterle (2015). In order to comply with the reduced class size, schools had to hire more teachers. Once the program was adopted on a national basis, the scarcity of teachers led to a decrease in the average quality of teachers. One can think of schools being forced not to fire bad teachers as there were simply no replacements available due to the high demand for teachers on a national level. Thus, the decreased quality of the pool of teachers led to unwanted policy effects which contradicted the original quasi-experimental research.

As the bilingual program expands including more and more schools, some general equilibrium effects might undermine the success of the program. As the demand for English speaking teachers increases, less experienced teachers qualified for English might displace older and otherwise better teachers. This example illustrates a situation where worse student performance in English-instructed content would arise not because of the increased difficulty of the instruction but due to the change in teacher quality.

In conclusion, this paper shows the estimates for a specific policy in a particular region of Spain. The potential effects of the same policy in a different region or a slightly different policy in the same region cannot be directly drawn from this study without taking into consideration new confounding factors. Nonetheless, this study sheds some light into the possible effects of this type of policy and shows results which can help design policy recommendations to improve the existing situation.

6.3 Policy Recommendations

Some of the limitations presented in Section 6.2 could have been avoided by following a different program implementation and better data collection.

6.3.1 Program Implementation

The implementation of the program could include a randomization wave conditional on observable characteristics for every school accepted into the program. The Regional Ministry of Education of Madrid includes newly qualified schools into the program yearly. Instead of simply accepting every qualified school into the bilingual program, 50% of the eligible schools could be accepted while the remaining 50% would be on hold until the next academic year. This way, suitable treatment and control groups could be formed by a researcher to estimate the causal effect of interest without the noise induced by student and school self-selection.

6.3.2 External Evaluations

Students in the bilingual program have their English skills assessed by external evaluations. Nevertheless, the external evaluation is only conducted in those schools participating in the bilingual program. The researcher needs to be provided with suitable control groups to properly measure the impact of the bilingual program on English. Otherwise, measuring the English improvement of students is not possible in a quasi-experimental context. Thus, an external evaluation for the English skills of those students not under the bilingual program is required to identify the causal effect of interest.

In Section 5.2 it was argued that the negative effect found in Science could be driven by a lack of Spanish vocabulary rather than a lack of content knowledge. The results provided evidence that the negative effect was not driven by the lack of Spanish vocabulary. Nevertheless, the question of *"Knowledge"* vis-a-vis *"Knowledge in Spanish"* could be better addressed with a well-designed standardized test adequate to distinguish between the two potential causes.

7 Robustness Checks

In this section, robustness checks for the results exposed in Section 5 are delivered. I will first focus on the average effect showing that the results hold under different treatment and control groups and alternative specifications. Secondly, I will provide an additional check for the distributional effects on Science.

7.1 Average Effect on Mathematics, Language and Science

In this section robustness checks corresponding to the average effect estimation corresponding to Section 4.3.1 are conducted.

7.1.1 Falsification Tests

In this section, the results from three different falsification tests are presented. These falsification tests will consist on running the same regressions with the same treatment and control groups but for years when the treatment schools were not yet under the bilingual program. Recall that the average treatment effect of the bilingual program on Science was found negative and significant effect. In Table 3 the effect estimated through equation 2 for the years 2013-2015 was displayed. The treatment schools were those schools which were under the bilingual program in 2015 but non-bilingual in 2013. Table 7 displays the estimation results of equation 2 for the same treatment and control schools but for the years 2011-2013, 2010-2011 and 2009-2010. During these time periods, the treatment schools were still not under the bilingual program. The estimated coefficient for the three tests is not significantly different from zero. This results provides evidence that the results obtained in Section 5 are not driven by special features of the specific years in which the analysis is conducted.

Table 7: Falsification Tests - Equation (2) Estimation Results

	Dependent Variable: Science		
Year	2011-2013	2010-2011	2009-2010
Diff-in-Diff	-0.065	-0.017	0.112
	(0.093)	(0.109)	(0.149)
Female	-0.277***	0.034**	-0.028
	(0.013)	(0.017)	(0.019)
Spanish	0.414***	0.839***	0.903***
	(0.031)	(0.036)	(0.030)
Age	-1.537***	-0.226***	-0.206***
	(0.024)	(0.061)	(0.054)
Constant	24.137***	8.814***	6.717***
	(0.297)	(0.753)	(0.662)
Observations	101,419	96,122	94,515
Adjusted R-squared	0.440	0.231	0.268

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

7.1.2 Excluding Repeaters

Students in bilingual schools with a low academic performance could systematically move to non-bilingual schools to avoid the increased difficulty of learning in a foreign language. In particular, students who decide to transfer to a different school when they need to retake an academic year due to low academic performance. Using the same specification given by equation (2) and the same treatment and control schools, I exclude students older than 12 from the analysis. The descriptive statistics for the modified groups can be found in 8. Table 9 displays the estimation results given by equation (2) excluding students older than 12. The findings show a negative and significant effect of the bilingual program on the performance of students in Science. The magnitude of this effect is very similar to the one obtained in Section 5.

Table 8: Descriptive Statistics - 2013 & 2015 - Excluding Repeaters

Mean / (SD)	2013		2015	
	Treatment	Control	Treatment	Control
Mathematics	7.353	7.122	7.392	7.147
	(2.074)	(2.161)	(2.385)	(2.474)
Language	8.247	8.125	7.857	7.796
	(1.623)	(1.698)	(1.599)	(1.722)
Science	8.460	8.311	6.315	6.430
	(1.487)	(1.577)	(2.234)	(2.260)
Female	0.483	0.506	0.486	0.500
	(0.500)	(0.500)	(0.500)	(0.500)
Spanish	0.948	0.923	0.966	0.940
	(0.223)	(0.266)	(0.182)	(0.237)
Observations	5,493	39,533	6,018	41,949

Table 9: Equation (2) Estimation Results for $t = \{2013, 2015\}$ Excluding Repeaters

Year 2013-2015	**Mathematics**	**Language**	**Science**
Diff-in-Diff	0.027	-0.065	-0.270***
	(0.080)	(0.062)	(0.095)
Female	-0.288***	0.143***	-0.312***
	(0.015)	(0.012)	(0.014)
Spanish	0.347***	0.437***	0.508***
	(0.037)	(0.031)	(0.032)
Constant	6.975***	7.663***	8.014***
	(0.038)	(0.031)	(0.035)
Observations	92,161	92,147	92,147
Adjusted R-squared	0.161	0.133	0.334

Standard Errors Clustered at the School level
* 10%, ** 5%, *** 1%

7.1.3 Using only bilingual schools

Non-bilingual schools could be intrinsically different from bilingual schools and therefore, not a good control group even when controlling for covariates. In this section, I use the same specification given by equation (2) and the same treatment schools. However, the control schools are only those non-bilingual schools that joined the program after 2010. Recall that when a school joins the bilingual program, it takes 5 years for its first bilingual cohort to sit the CDI exam in 6th grade. For example, the bilingual schools in 2015 are the bilingual schools with a bilingual cohort in the 2015 CDI exam. The descriptive statistics for the modified groups can be found in 10. Table 11 displays the estimation results given by equation (2) using only future bilingual schools as controls. The findings show a negative and significant effect of the bilingual program on the performance of students in Science. The magnitude of this effect is very similar to the one obtained in Section 5.

Table 10: Descriptive Statistics $t = \{2013, 2015\}$ Using only bilingual schools

Mean / (SD)	2013		2015	
	Treatment	Control	Treatment	Control
Mathematics	7.097	6.878	7.179	6.959
	(2.260)	(2.325)	(2.521)	(2.581)
Language	8.032	7.937	7.707	7.633
	(1.842)	(1.848)	(1.741)	(1.843)
Science	8.273	8.162	6.141	6.236
	(1.658)	(1.691)	(2.316)	(2.338)
Female	0.475	0.498	0.482	0.499
	(0.499)	(0.500)	(0.500)	(0.500)
Spanish	0.931	0.913	0.955	0.929
	(0.254)	(0.282)	(0.208)	(0.258)
Age	12.099	12.124	12.081	12.107
	(0.309)	(0.344)	(0.294)	(0.325)
Observations	6,128	11,787	6,618	12,930

Table 11: Equation (2) Estimation Results for $t = \{2013, 2015\}$ Using only bilingual schools

Year 2013-2015	**Mathematics**	**Language**	**Science**
Diff-in-Diff	-0.002	-0.044	-0.234**
	(0.095)	(0.071)	(0.112)
Female	-0.292***	0.178***	-0.301***
	(0.025)	(0.020)	(0.023)
Spanish	0.221***	0.405***	0.440***
	(0.060)	(0.056)	(0.053)
Age	-2.070***	-1.621***	-1.616***
	(0.051)	(0.041)	(0.040)
Constant	31.967***	27.144***	27.510***
	(0.628)	(0.495)	(0.491)
Observations	36,797	36,795	36,795
Adjusted R-squared	0.223	0.207	0.381

Standard Errors Clustered at the School level
* 10%, ** 5%, *** 1%

7.1.4 Same Model using data from 2011 and 2013

In this section, the estimation results of the same model used in Section 4.3.1 are presented but using data from 2011 and 2013. With this test, we want to see if the estimated effect corresponds only to the very specific treatment schools considered in the previous model. The treatment schools will be those schools under the bilingual program in 2013 but which were non-bilingual in 2011. The control schools will be the rest of non-bilingual schools in 2013. The descriptive statistics for the modified groups can be found in 2. Table 13 displays the estimation results given by equation (2). The findings show a negative and significant effect of the bilingual program on the performance of students in Science. The magnitude of this effect is very similar to the one obtained in Section 5.

Table 12: Descriptive Statistics - 2011 & 2013

Mean / (SD)	2011 Treatment	2011 Control	2013 Treatment	2013 Control
Mathematics	5.828	5.903	6.799	6.800
	(3.019)	(2.943)	(2.403)	(2.386)
Language	7.409	7.570	7.834	7.830
	(2.541)	(2.510)	(1.921)	(1.969)
Science	5.567	5.525	7.890	8.061
	(2.605)	(2.496)	(1.858)	(1.795)
Female	0.481	0.490	0.489	0.492
	(0.500)	(0.500)	(0.500)	(0.500)
Spanish	0.900	0.926	0.882	0.897
	(0.300)	(0.262)	(0.323)	(0.304)
Age	12.188	12.150	12.139	12.140
	(0.431)	(0.382)	(0.357)	(0.359)
Observations	2,804	52,003	2,838	52,716

Table 13: Estimation Results of Equation (2) for $t = \{2011, 2013\}$

Year 2011-2013	**Mathematics**	**Language**	**Science**
Diff-in-Diff	0.023	0.112	-0.259**
	(0.133)	(0.091)	(0.127)
Female	-0.289***	0.257***	-0.281***
	(0.015)	(0.012)	(0.012)
Spanish	0.111***	0.272***	0.433***
	(0.036)	(0.031)	(0.031)
Age	-2.065***	-1.821***	-1.541***
	(0.029)	(0.026)	(0.023)
Constant	31.226***	29.555***	24.177***
	(0.362)	(0.313)	(0.286)
Observations	106,906	106,913	106,913
Adjusted R-squared	0.275	0.258	0.437

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

7.1.5 Alternative Specification

In this section, a twofold robustness check is conducted. On the one hand, an alternative specification to to the one given by equation (2) is provided. Additionally, the estimation results of this specification for the years 2013 and 2011 is presented. The specification chosen is the one used by Anghel, Cabrales and Carro (2015) in their evaluation of the bilingual program of Madrid for the years 2009-2010 and 2010-2011. The treatment group is constituted by the bilingual schools that were bilingual in 2015 but not in 2013. The control group is constituted by the non-bilingual schools. The descriptive statistics for treatment and control schools can be found in Tables (12) and (2) for the years $t = \{2011, 2013\}$ and $t = \{2013, 2015\}$ respectively. Consider the new specification:

$$Y_{ijt} = \beta_0 + \beta_1 MCB_j + \beta_2 D_t + \tau_{DID}(D_t \times MCB_j) + \delta' X_{ijt} + u_{ijt} \tag{5}$$

Where Y_{ijt} is the CDI exam grade, MCB_j is a treatment school dummy, D_t is a year dummy and X_{ijt} are student-level covariates. Tables (14) and (15) display the estimation results for equation (5) for the years $t = \{2011, 2013\}$ and $t = \{2013, 2015\}$ respectively.

The findings show a negative and significant effect of the bilingual program on the performance of students in Science. The magnitude of this effect is very similar to the one obtained in Section 5.

Table 14: Equation (5) Estimation Results for $t = \{2011, 2013\}$

Year 2011-2013	**Mathematics**	**Language**	**Science**
Diff-in-Diff	-0.011	0.090	-0.288**
	(0.136)	(0.089)	(0.125)
Bilingual School Dummy	0.017	-0.074	0.126
	(0.150)	(0.090)	(0.132)
Year Dummy	0.700***	0.013	2.360***
	(0.031)	(0.022)	(0.030)
Female	-0.292***	0.261***	-0.269***
	(0.020)	(0.015)	(0.017)
Spanish	0.511***	0.604***	0.677***
	(0.047)	(0.038)	(0.039)
Age	-2.353***	-2.041***	-1.741***
	(0.035)	(0.030)	(0.027)
Constant	34.348***	31.927***	26.364***
	(0.428)	(0.366)	(0.337)
Observations	106,906	106,913	106,913
R-squared	0.142	0.161	0.334

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

Table 15: Equation (5) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	**Mathematics**	**Language**	**Science**
Diff-in-Diff	0.035	-0.061	-0.237**
	(0.083)	(0.065)	(0.098)
Bilingual School Dummy	0.200**	0.122**	0.123*
	(0.088)	(0.062)	(0.065)
Year Dummy	-0.003	-0.320***	-1.946***
	(0.031)	(0.021)	(0.032)
Female	-0.289***	0.185***	-0.306***
	(0.021)	(0.015)	(0.018)
Spanish	0.436***	0.557***	0.564***
	(0.042)	(0.034)	(0.036)
Age	-2.377***	-1.925***	-1.906***
	(0.034)	(0.030)	(0.029)
Constant	35.380***	30.594***	30.832***
	(0.417)	(0.366)	(0.352)
Observations	105,684	105,668	105,668
R-squared	0.120	0.147	0.269

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

7.2 Distributional Effects

In this section, robustness checks for the distributional effect estimation corresponding to Section 4.3.3 are conducted.

7.2.1 Distributional Effects on Mathematics and Language

In this section, the distributional effects estimated for Mathematics and Language are presented. Throughout the paper, these two Spanish-instructed subjects served as controls for the effect estimated in Science, English-instructed. The distributional effects estimated for Mathematics and Language can be found in Tables 16 and 17 respectively. Section 8.2 displays Tables 22 and 23 which are the extended version of Tables 16 and 17 including more quantiles. None of the estimated coefficients are significantly different from zero for those subjects instructed in Spanish. A graphical representation of the estimated distributional effects for Mathematics and Language can be found in Figures (2) and (3) respectively.

Table 16: Equation (4) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Mathematics $\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Diff-in-Diff	-1.605	-0.741	-1.077	-4.502	-5.957
	(5.109)	(5.098)	(4.501)	(4.758)	(5.072)
Female	5.420	-0.523	1.275	4.858	4.268
	(12.262)	(11.497)	(10.968)	(10.813)	(11.703)
Spanish	13.777	30.267**	32.728**	32.643**	40.081***
	(16.433)	(14.651)	(14.151)	(14.544)	(15.178)
Age	55.072***	41.062***	16.629	15.098	-4.280
	(14.465)	(13.274)	(13.014)	(13.085)	(14.481)
Constant	-660.030***	-493.419***	-198.302	-182.364	47.892
	(179.637)	(165.923)	(162.223)	(164.044)	(181.329)
Observations	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.086	0.108	0.105	0.100	0.127

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

Table 17: Equation (4) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Language $\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Diff-in-Diff	0.201	-2.230	-4.159	-1.475	-2.585
	(5.170)	(4.815)	(3.912)	(3.678)	(4.108)
Female	-3.758	-4.761	-6.679	-4.356	3.230
	(11.962)	(10.418)	(9.331)	(8.807)	(9.857)
Spanish	16.154	23.590*	14.080	12.518	8.392
	(16.317)	(13.756)	(11.431)	(10.681)	(10.913)
Age	54.737***	23.446*	7.454	4.729	-19.516
	(17.247)	(13.136)	(12.387)	(11.127)	(12.369)
Constant	-651.663***	-272.398*	-69.238	-38.045	255.733*
	(212.091)	(161.003)	(153.533)	(138.183)	(153.099)
Observations	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.102	0.086	0.071	0.068	0.051

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

7.3 Distributional Effects - Non-Linear Model

In this Section, a modified version of the model utilized to estimate the distributional effects is presented. The distributional analysis is addressing the heterogeneous effect of the policy across the distribution of student ability. However, while the effect could be argued to be linear around the median, it is likely to be non-linear at the tails of the distribution. Namely, as we are measuring the impact of the increased difficulty on students performance, we expect the effect not to be properly captured by a regular 0 to 10 scale because low performing children cannot obtain a grade below zero!

7.3.1 Understanding the Model Differently

In Section 7.2, the 2 step methodology to estimate the distributional effects was introduced. The second step was understood to provide the effect of the school-level policy on the quantile net of observable student-level characteristics. In other words, the parameter of interest was capturing the variation in student grades which could not be explained by student individual characteristics at each school. As the dependent variable was not simply an exam grade, the estimated effect provides point estimates hard to understand beyond the back of the envelope calculation mentioned. However, we can also see that the model implied by the 2 step methodology is the following:

$$\mathrm{Q}_\tau y_{ist} = \beta'_{st}(\tau)x_{ist} + \eta_s(\tau) + \theta_t(\tau) + \psi(\tau)(D_t \times MCB_s) + \delta'(\tau)X_{st} + u_{ist}(\tau) \tag{6}$$

Where x_{ist} are student level covariates, η_s is the school fixed effect, $\theta_t(\tau)$ is the year fixed effect, MCB_s is a treatment school dummy, D_t is a year dummy and X_{st} are observable school characteristics. $\psi(\tau)$ is the parameter of interest measuring the impact of the bilingual policy on students' grades for each quantile τ.

As only x_{ist} and the dependent variable are on a student level, the first step consisted on estimating $\hat{\beta}'_{st}$ using the within school distribution of students. Then, the second step consisted in running a school level regression to estimate the effect of the policy for each student quantile. More precisely, $\hat{\alpha}_s(\tau)$ was regressed on school level characteristics to capture the effect of the policy on student performance for the different quantiles net of the individual covariates[15]. If we clear for $\mathrm{Q}^s_{y_{ist}}$ we have the implied model given by equation 6.

7.3.2 Model with transformed dependent variable

So far, the grade obtained in Science could take any value from 0 to 10. The transformation applied to this test score is the following:

$$\tilde{y}_{ist} = \log\left(\frac{y_{ist}}{10 - y_{ist}}\right) \iff y_{ist} = 10 \cdot \frac{e^{\tilde{y}_{ist}}}{1 + e^{\tilde{y}_{ist}}} \iff y_{ist} = 10 \cdot \Lambda(\tilde{y}_{ist})$$

Step 1: As in the previous model, quantile regressions within each school are performed to estimate the effect of micro-level covariates on individual student outcomes.

$$\mathrm{Q}^s_\tau(\tilde{y}_{it}) = \beta'^s(\tau)x_{it} + \alpha^s(\tau) + \epsilon(\tau) \tag{7}$$

[15]Where $\hat{\alpha}_s(\tau) = \frac{1}{N_s}\sum_i \left(\mathrm{Q}_\tau y_{ist} - \beta'_{st}(\tau)x_{ist}\right)$. N_s is the number of students in school s

Where $Q_\tau^s(\tilde{y}_{it})$ is the τth conditional quantile of $\tilde{y}_{it}$ in school s, $\tilde{y}_{it}$ is the transformed student grade in school s which is given by the CDI grade in Mathematics, Language or Science. x_{it} are student level covariates including age and dummies for gender and Spanish nationality.

Step 2: As in the previous model, $\hat{\alpha}^s(\tau)$ is regressed on school level covariates. The regression estimated in step 2 is given by:

$$\hat{\alpha}^s(\tau) = \eta_s(\tau) + \theta_t(\tau) + \psi(\tau)(D_t \times MCB_s) + \delta'(\tau)X_{st} + u_{ist}(\tau)$$

Which is equivalent to:

$$Q_\tau \tilde{y}_{ist} = \beta'_{st}(\tau)\, x_{ist} + \eta_s(\tau) + \theta_t(\tau) + \psi(\tau)(D_t \times MCB_s) + \delta'(\tau)X_{st} + u_{ist}(\tau) \tag{8}$$

Where x_{ist} are student level covariates, η_s is the school fixed effect, $\theta_t(\tau)$ is the year fixed effect, MCB_s is a treatment school dummy, D_t is a year dummy and X_{st} are observable school characteristics. $\psi(\tau)$ is the parameter of interest measuring the impact of the bilingual policy on students' transformed grades for each quantile τ. $Q_\tau \tilde{y}_{ist}$ is the τth quantile of $\tilde{y}_{ist}$, the transformed student grade.

Provided that the quantile function allows us to do the following:

$$\log\left(\frac{Q_\tau(z)}{10 - Q_\tau(z)}\right) = Q_\tau\left(\log\left(\frac{z}{10 - z}\right)\right) \tag{9}$$

. It can be shown that:

$$Q_\tau(y_{ist}) = 10 \cdot \Lambda(Q_\tau(\tilde{y}_{ist}))$$

Step 3: To recover the impact of the policy on the original grade scale we can rewrite equation 8 in terms of the original dependent variable with the following equation:

$$Q_\tau y_{ist} = 10 \cdot \Lambda\Big(\beta'_{st}(\tau)x_{ist} + \eta_s(\tau) + \theta_t(\tau) + \psi(\tau)(D_t \times MCB_s) + \delta'(\tau)X_{st} + u_{ist}(\tau)\Big) \tag{10}$$

The impact of the policy on the original variable is given by:

$$\frac{\partial Q_\tau(y_s)}{\partial DiD} = \frac{\partial\big(10 \cdot \Lambda(Q_\tau(\tilde{y}_s))\big)}{\partial DiD} = 10 \cdot \psi(\tau)\frac{1}{S}\sum_s \Lambda'(Q_\tau(\tilde{y}_s)) \tag{11}$$

Where S is the number of schools and $\psi(\tau)$ is the parameter of interest estimated by equation 8. In other words, $\psi(\tau)$ is the effect of the policy on the transformed variable $\tilde{y}_s$

7.3.3 Results

In this section, the estimation results of the model described in Section 7.3.2 are presented. First, the estimation results corresponding to step 2 will be provided. Then, the estimation results corresponding to equation 11 will be presented.

Step 2 - Estimation Results

The descriptive statistics for the treatment and control groups can be found in Table (2). Table (18) shows the estimation results for the years 2013-2015. The estimation of the parameter of interest $\psi(\tau)$ capturing the impact of the bilingual program on the transformed variable of Science is found to be negative and significant for all quantiles but 0.1. An extended version of these results, including all deciles, can be found in Table 24.

Table 18: Equation (8) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Science				
	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Diff-in-Diff	-1.110	-4.515***	-5.386***	-4.191**	-5.989**
	(1.888)	(1.476)	(1.435)	(1.805)	(2.479)
Female	1.384	-1.504	-5.731*	-7.100*	0.800
	(4.702)	(3.722)	(3.464)	(4.195)	(5.774)
Spanish	10.299	15.882***	16.045***	12.900**	13.776**
	(6.474)	(5.042)	(5.151)	(6.040)	(6.474)
Age	28.275***	11.529*	9.414*	5.798	-3.930
	(6.739)	(5.920)	(5.083)	(6.064)	(7.602)
Constant	-344.473***	-141.845*	-112.047*	-61.045	57.953
	(83.839)	(73.382)	(63.178)	(75.784)	(94.976)
Observations	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.151	0.067	0.049	0.025	0.056

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

Step 3 - Estimation Results

The estimation results for equation 11 are displayed in Table 19. The estimated effects are negative and significantly for all quantiles but 0.1. The findings show a negative and significant effect of the bilingual program on the performance of students in Science. The magnitude of this effect is very similar to the one obtained in Section 5.

Table 19: Equation (8) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Science				
	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Diff-in-Diff	-2.316	-9.073***	-9.265***	-5.495***	-4.600*
	(1.888)	(1.476)	(1.435)	(1.805)	(2.479)
Observations	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.151	0.067	0.049	0.025	0.056

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

7.3.4 Distributional Effects on Science - Alternative Specification

In this section, an alternative specification for the second step of the methodology presented in Section 4.3.3 is shown. Consider the alternative specification:

$$\alpha_{st}(\tau) = \beta_0(\tau) + \beta_1(\tau)D_t + \beta_2(\tau)MCB_s + \psi(\tau)(D_t \times MCB_s) + \delta'(\tau)X_{st} + u_{st}(\tau) \quad (12)$$

Where MCB_s is a Treatment School Dummy, D_t is a Year Dummy and X_{st} are observable school characteristics. $\psi(\tau)$ is the parameter of interest measuring the impact of the bilingual policy on students' grades for each quantile τ.

Table (20) shows the estimation results of equation (12). The negative impact of the bilingual program on the students' performance on Science is concentrated around the median while no significant effect is found in the tails of the distribution. The findings show a negative and significant effect of the bilingual program on the performance of students in Science. The magnitude of this effect is very similar to the one obtained in Section 5.

Table 20: Equation (12) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Science				
	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Diff-in-Diff	-2.734	-7.294**	-8.311***	-4.153	-3.105
	(3.324)	(2.879)	(2.477)	(2.676)	(2.934)
Bilingual School Dummy	0.721	3.848**	3.704**	2.607	2.185
	(2.103)	(1.803)	(1.525)	(1.609)	(1.589)
Year Dummy	-6.810***	-2.404***	1.265	4.070***	7.451***
	(0.949)	(0.829)	(0.777)	(0.819)	(0.921)
Female	-3.798	-6.081	-3.963	-4.201	3.086
	(4.136)	(3.894)	(3.895)	(3.939)	(4.538)
Spanish	9.588*	12.369***	9.158**	8.010*	6.148
	(5.275)	(4.454)	(4.185)	(4.118)	(4.144)
Age	39.601***	17.557***	12.448**	5.388	-13.146***
	(6.772)	(5.487)	(5.380)	(4.838)	(4.711)
Constant	-464.690***	-194.342***	-130.999*	-44.747	178.278***
	(85.810)	(69.658)	(68.143)	(61.439)	(59.610)
Observations	2,247	2,247	2,247	2,247	2,247
R-squared	0.067	0.019	0.011	0.015	0.044

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

8 Appendix

8.1 Figures

Figure 1: Equation (4) Science Estimation Results for $t = \{2013, 2015\}$

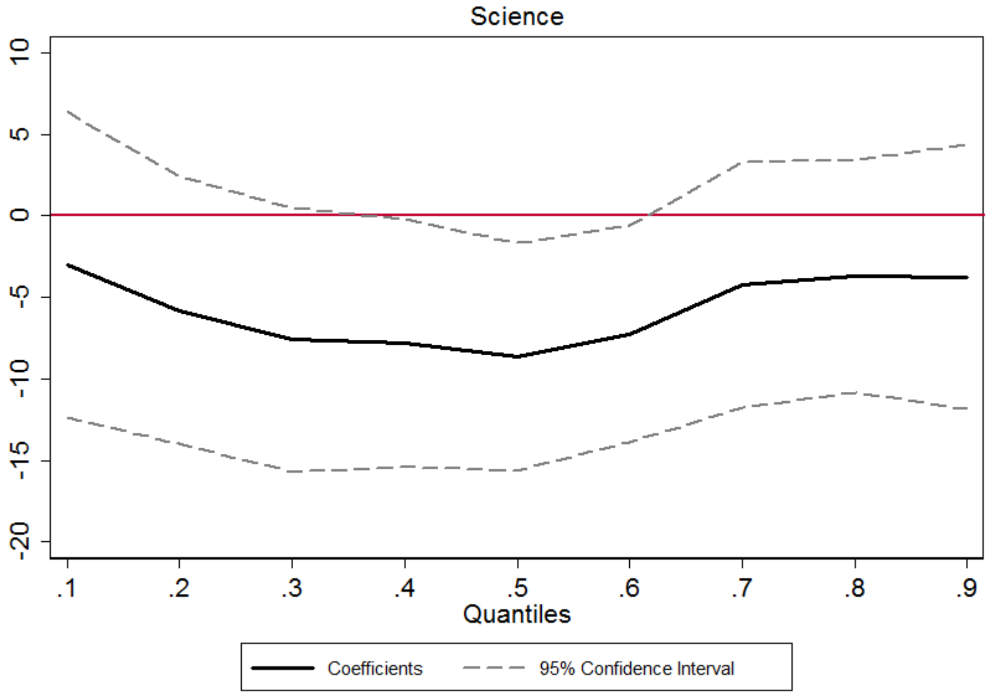


Figure 2: Equation (4) Mathematics Estimation Results for $t = \{2013, 2015\}$

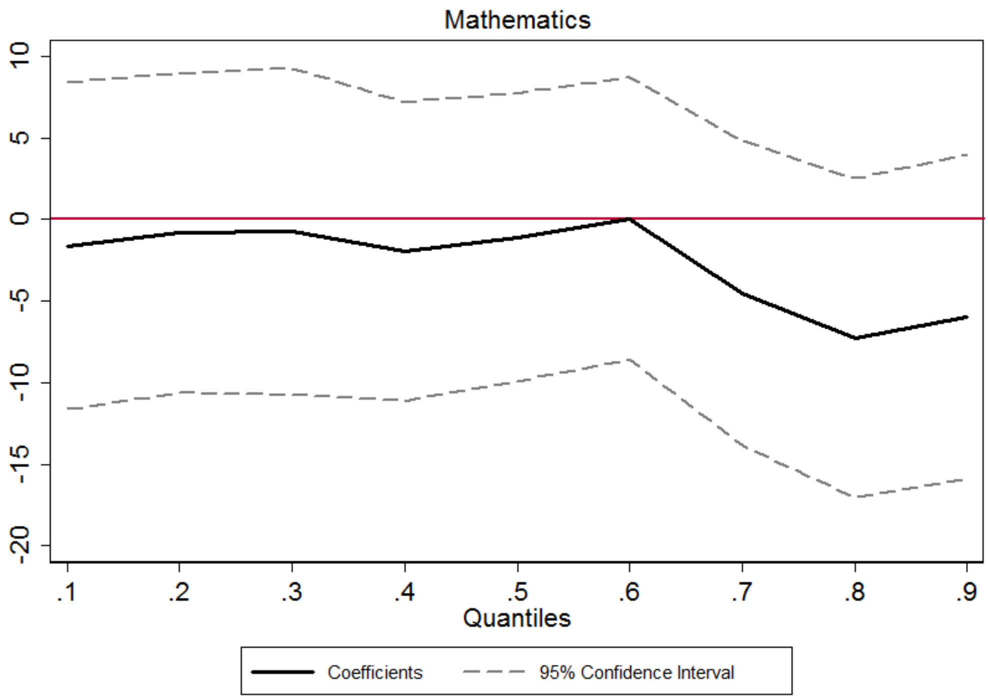
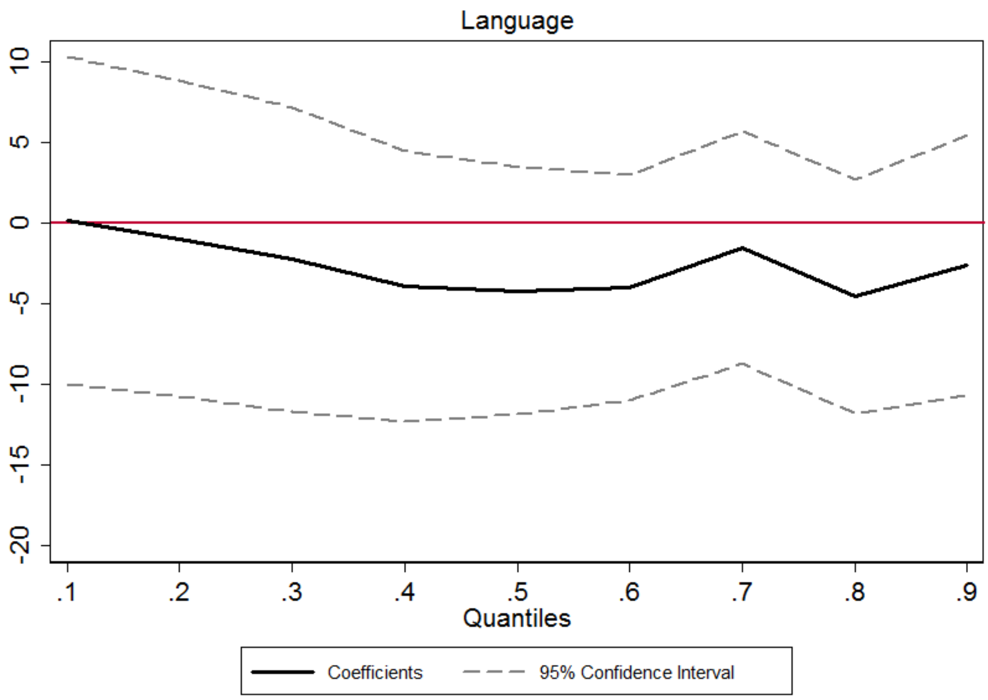


Figure 3: Equation (4) Language Estimation Results for $t = \{2013, 2015\}$



8.2 Equation (4) Estimation Results for $t = \{2013, 2015\}$

Table 21: Equation (4) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Science								
	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
Diff-in-Diff	-2.987	-5.783	-7.606*	-7.807**	-8.650**	-7.232**	-4.232	-3.687	-3.742
	(4.777)	(4.183)	(4.127)	(3.876)	(3.555)	(3.379)	(3.845)	(3.628)	(4.139)
Female	1.725	-1.573	-7.483	-7.144	-4.106	-5.622	-9.599	-1.947	9.337
	(10.600)	(9.285)	(9.091)	(8.946)	(8.522)	(8.288)	(9.158)	(9.819)	(10.821)
Spanish	13.962	22.453*	29.566**	29.959***	28.287***	30.649***	27.234**	26.815**	24.723**
	(14.576)	(12.692)	(12.173)	(10.774)	(10.356)	(11.580)	(12.684)	(11.836)	(12.505)
Constant	-657.986***	-438.731***	-304.749*	-248.657	-195.356	-254.471*	-188.293	-40.864	80.324
	(181.683)	(165.094)	(158.725)	(151.869)	(153.888)	(148.809)	(163.457)	(155.829)	(188.381)
Observations	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.147	0.111	0.080	0.074	0.093	0.099	0.062	0.078	0.090

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

Table 22: Equation (4) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Mathematics								
	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
Diff-in-Diff	-1.605	-0.829	-0.741	-1.950	-1.077	0.072	-4.502	-7.268	-5.957
	(5.109)	(4.991)	(5.098)	(4.669)	(4.501)	(4.407)	(4.758)	(4.983)	(5.072)
Female	5.420	-1.176	-0.523	0.125	1.275	4.385	4.858	2.220	4.268
	(12.262)	(11.431)	(11.497)	(10.771)	(10.968)	(11.325)	(10.813)	(11.218)	(11.703)
Spanish	13.777	24.950	30.267**	35.172**	32.728**	33.374**	32.643**	39.346***	40.081***
	(16.433)	(15.341)	(14.651)	(14.389)	(14.151)	(14.166)	(14.544)	(14.871)	(15.178)
Age	55.072***	42.836***	41.062***	30.322**	16.629	16.130	15.098	11.161	-4.280
	(14.465)	(12.559)	(13.274)	(13.198)	(13.014)	(13.456)	(13.085)	(13.872)	(14.481)
Constant	-660.030***	-513.288***	-493.419***	-366.683**	-198.302	-194.253	-182.364	-138.865	47.892
	(179.637)	(156.811)	(165.923)	(164.482)	(162.223)	(168.707)	(164.044)	(174.399)	(181.329)
Observations	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.086	0.118	0.108	0.106	0.105	0.081	0.100	0.114	0.127

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

Table 23: Equation (4) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Language								
	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
Diff-in-Diff	0.201	-0.939	-2.230	-3.892	-4.159	-3.974	-1.475	-4.521	-2.585
	(5.170)	(4.988)	(4.815)	(4.276)	(3.912)	(3.559)	(3.678)	(3.699)	(4.108)
Female	-3.758	0.100	-4.761	-5.316	-6.679	-6.499	-4.356	-8.359	3.230
	(11.962)	(10.226)	(10.418)	(8.925)	(9.331)	(8.808)	(8.807)	(9.289)	(9.857)
Spanish	16.154	15.135	23.590*	19.759*	14.080	18.507*	12.518	8.479	8.392
	(16.317)	(14.483)	(13.756)	(11.054)	(11.431)	(10.758)	(10.681)	(9.998)	(10.913)
Age	54.737***	33.384**	23.446*	13.668	7.454	3.846	4.729	-11.477	-19.516
	(17.247)	(13.472)	(13.136)	(12.347)	(12.387)	(10.872)	(11.127)	(11.458)	(12.369)
Constant	-651.663***	-390.267**	-272.398*	-150.175	-69.238	-30.668	-38.045	163.725	255.733*
	(212.091)	(165.202)	(161.003)	(152.728)	(153.533)	(135.540)	(138.183)	(141.867)	(153.099)
Observations	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.102	0.052	0.086	0.099	0.071	0.089	0.068	0.048	0.051

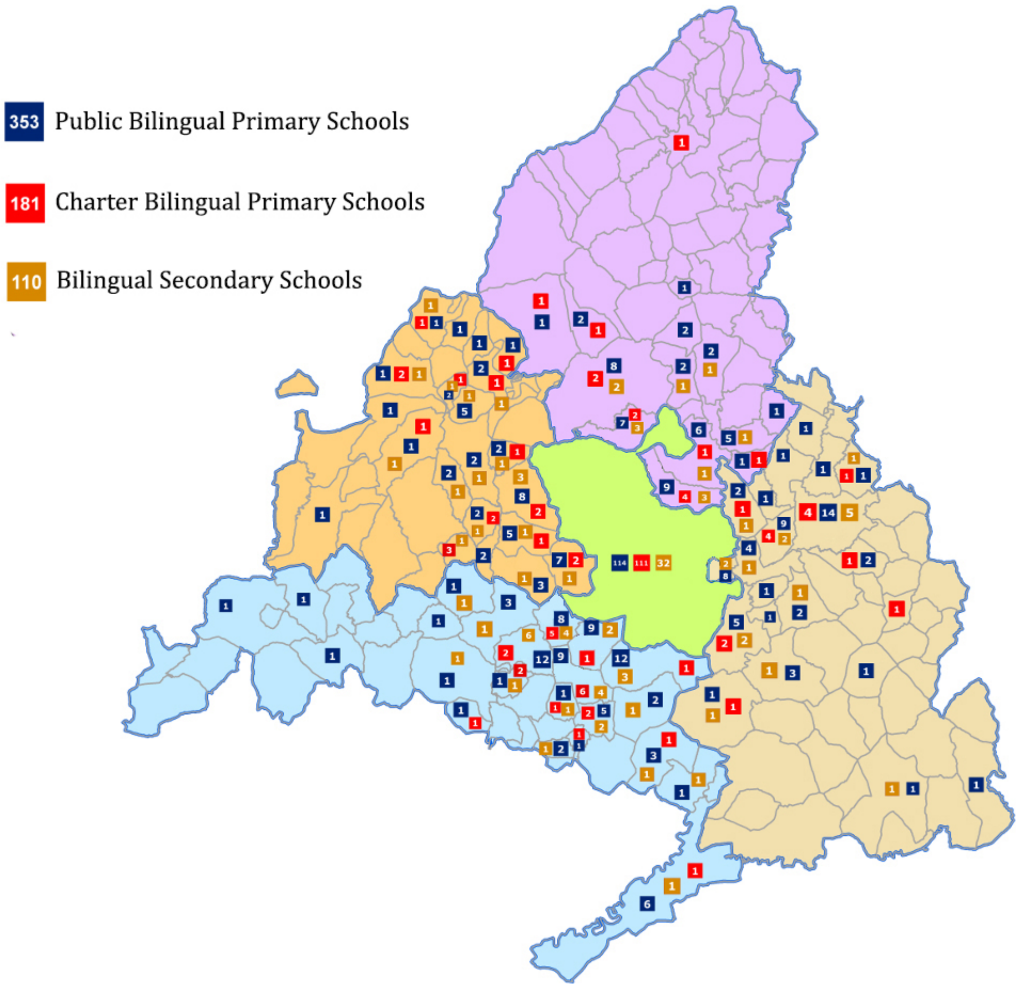
Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

Table 24: Equation (8) Estimation Results for $t = \{2013, 2015\}$

Year 2013-2015	Dependent Variable: Science								
	$\tau = 0.1$	$\tau = 0.2$	$\tau = 0.3$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.7$	$\tau = 0.8$	$\tau = 0.9$
Diff-in-Diff	-1.110	-3.091**	-4.515***	-5.086***	-5.386***	-6.057***	-4.191**	-5.894***	-5.989**
	(1.888)	(1.544)	(1.476)	(1.441)	(1.435)	(1.462)	(1.805)	(1.915)	(2.479)
Female	1.384	0.409	-1.504	-1.907	-5.731*	-7.866**	-7.100*	-1.207	0.800
	(4.702)	(3.740)	(3.722)	(3.451)	(3.464)	(3.560)	(4.195)	(4.915)	(5.774)
Spanish	10.299	11.937**	15.882***	14.534***	16.045***	14.377***	12.900**	20.110***	13.776**
	(6.474)	(5.332)	(5.042)	(4.906)	(5.151)	(5.517)	(6.040)	(6.143)	(6.474)
Age	28.275***	16.348***	11.529*	10.511*	9.414*	6.810	5.798	4.326	-3.930
	(6.739)	(5.958)	(5.920)	(5.658)	(5.083)	(5.214)	(6.064)	(6.559)	(7.602)
Constant	-344.473***	-199.032***	-141.845*	-127.242*	-112.047*	-76.348	-61.045	-49.235	57.953
	(83.839)	(73.852)	(73.382)	(70.387)	(63.178)	(64.859)	(75.784)	(82.164)	(94.976)
Observations	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247	2,247
Adjusted R-squared	0.151	0.095	0.067	0.076	0.049	0.057	0.025	0.074	0.056

Standard Errors Clustered at School level
* 10%, ** 5%, *** 1%

8.3 Geographical Distribution of Bilingual Schools in Madrid



References

Admiraal, Wilfried, Gerard Westhoff, and Kees de Bot. 2006. "Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English 1." *Educational Research and Evaluation*, 12(1): 75–93.

Anghel, Brindusa, Antonio Cabrales, and Jesus M Carro. 2015. "Evaluating a bilingual education program in Spain: the impact beyond foreign language learning." *Economic Inquiry.*

Angrist, Joshua D, and Victor Lavy. 1997. "Using Maimonides' rule to estimate the effect of class size on student achievement." National Bureau of Economic Research.

Appel, René, and Pieter Muysken. 2006. *Language contact and bilingualism.* Amsterdam University Press.

Bietenbeck, Jan. 2014. "Teaching practices and cognitive skills." *Labour Economics*, 30: 143–153.

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

Chetverikov, Denis, Bradley Larsen, and Christopher Palmer. 2016. "IV Quantile Regression for Group-level Treatments, with an Application to the Effects of Trade on the Distribution of Wages." *Econometrica, vol 84, No.2.*

Christian, Donna. 1994. "Two-Way Bilingual Education: Students Learning through Two Languages. Educational Practice Report: 12."

Cromdal, Jakob. 1999. "Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish–English bilinguals." *Applied Psycholinguistics*, 20(01): 1–20.

Cummins, Jim. 2000. "Immersion education for the millennium: What we have learned from 30 years of research on second language immersion." *Retrieved April*, 16: 2006.

Diaz, Rafael M. 1985. "Bilingual cognitive development: Addressing three gaps in current research." *Child development*, 1376–1388.

Dieterle, Steven G. 2015. "Class-size reduction policies and the quality of entering teachers." *Labour Economics*, 36: 35–47.

European Union Council. 2008. "European Strategy for Multilingualism." *Journal of the European Union Council Resolution.*

Fidrmuc, Jan, and Jarko Fidrmuc. 2009. "Foreign languages and trade."

Genesee, Fred. 1987. *Learning through two languages: Studies of immersion and bilingual education.* Newbury house publishers.

Gerena, Linda, and María Dolores Ramírez Verdugo. 2014. "Analyzing Bilingual Teaching and Learning in Madrid, Spain: A Fulbright Scholar Collaborative Research Project." *Gist: Education and Learning Research Journal*, , (8): 118–136.

Ginsburgh, Victor A, and Juan Prieto-Rodriguez. 2011. "Returns to foreign languages of native workers in the European Union." *Industrial & Labor Relations Review*, 64(3): 599–618.

Goldhaber, Dan D, and Dominic J Brewer. 1997. "Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity." *Journal of Human Resources*, 505–523.

Gonzalez, Arturo. 2000. "The acquisition and labor market value of four English skills: new evidence from NALS." *Contemporary Economic Policy*, 18(3): 259–269.

Greene, Jay Phillip. 1998. *A meta-analysis of the effectiveness of bilingual education.* Tomas Rivera Policy Institute Claremont, CA.

Hausman, Jerry A, and William E Taylor. 1981. "Panel data and unobservable individual effects." *Econometrica: Journal of the Econometric Society*, 1377–1398.

Heckman, James J, and Edward Vytlacil. 2005. "Structural equations, treatment effects, and econometric policy evaluation1." *Econometrica*, 73(3): 669–738.

Jepsen, Christopher. 2010. "Bilingual education and English proficiency." *Education*, 5(2): 200–227.

Krueger, Alan B. 2003. "Economic considerations and class size*." *The Economic Journal*, 113(485): F34–F63.

Krueger, Alan B, and Diane M Whitmore. 2001. "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR." *The Economic Journal*, 111(468): 1–28.

Lavy, Victor. 2015. "What makes an effective teacher? Quasi-experimental evidence." *CESifo Economic Studies*, ifv001.

Lucas, Robert E. 1976. "Econometric policy evaluation: A critique." Vol. 1, 19–46, Elsevier.

Murnane, Richard J, and Barbara R Phillips. 1981. "What do effective teachers of inner-city children have in common?" *Social Science Research*, 10(1): 83–100.

Schwerdt, Guido, and Amelie C Wuppermann. 2011. "Is traditional teaching really all that bad? A within-student between-subject approach." *Economics of Education Review*, 30(2): 365–379.

Silaghi, Florina Raluca. 2011. "Immigration and peer effects: evidence from primary education in Spain." PhD diss. Master thesis CEMFI.

Trejo, Stephen J. 2003. "Intergenerational progress of Mexican-origin workers in the US labor market." *Journal of Human Resources*, 38(3): 467–489.

MASTER'S THESIS CEMFI

0801 *Paula Inés Papp*: "Bank lending in developing countries: The effects of foreign banks".

0802 *Liliana Bara*: "Money demand and adoption of financial technologies: An analysis with household data".

0803 *J. David Fernández* Fernández: "Elección de cartera de los hogares españoles: El papel de la vivienda y los costes de participación".

0804 *Máximo Ferrando Ortí*: "Expropriation risk and corporate debt pricing: the case of leveraged buyouts".

0805 *Roberto Ramos*: "Do IMF Programmes stabilize the Economy?".

0806 *Francisco Javier Montenegro*: "Distorsiones de Basilea II en un contexto multifactorial".

0807 *Clara Ruiz Prada*: "Do we really want to know? Private incentives and the social value of information".

0808 *Jose Antonio Espin*: "The "bird in the hand" is not a fallacy: A model of dividends based on hidden savings".

0901 *Víctor Capdevila Cascante*: "On the relationship between risk and expected return in the Spanish stock market".

0902 *Lola Morales*: "Mean-variance efficiency tests with conditioning information: A comparison".

0903 *Cristina Soria Ruiz-Ogarrio*: "La elasticidad micro y macro de la oferta laboral familiar: Evidencia para España".

0904 *Carla Zambrano Barbery*: "Determinants for out-migration of foreign-born in Spain".

0905 *Álvaro de Santos Moreno*: "Stock lending, short selling and market returns: The Spanish market".

0906 *Olivia Peraita*: "Assessing the impact of macroeconomic cycles on losses of CDO tranches".

0907 *Iván A. Kataryniuk Di Costanzo:* "A behavioral explanation for the IPO puzzles".

1001 *Oriol Carreras:* "Banks in a dynamic general equilibrium model".

1002 *Santiago Pereda-Fernández:* "Quantile regression discontinuity: Estimating the effect of class size on scholastic achievement".

1003 *Ruxandra Ciupagea:* "Competition and "blinders": A duopoly model of information provision".

1004 *Rebeca Anguren:* "Credit cycles: Evidence based on a non-linear model for developed countries".

1005 *Alba Diz:* "The dynamics of body fat and wages".

1101 *Daniela Scidá:* "The dynamics of trust: Adjustment in individual trust levels to changes in social environment".

1102 *Catalina Campillo:* "Female labor force participation and mortgage debt".

1103 *Florina Raluca Silaghi:* "Immigration and peer effects: Evidence from primary education in Spain".

1104 *Jan-Christoph Bietenbeck:* "Teaching practices and student achievement: Evidence from TIMSS".

1105 *Andrés Gago:* "Reciprocity: Is it outcomes or intentions? A laboratory experiment".

1201 *Rocío Madera Holgado:* "Dual labor markets and productivity".

1202 *Lucas Gortazar:* "Broadcasting rights in football leagues and TV competition".

1203 *Mª. Elena Álvarez Corral:* "Gender differences in labor market performance: Evidence from Spanish notaries".

1204 *José Alonso Olmedo:* "A political economy approach to banking regulation".

1301 *Joaquín García-Cabo Herrero:* "Unemployment and productivity over the business cycle: Evidence from OECD countries".

1302 *Luis Díez Catalán:* "Collective bargaining and unemployment during the Great Recession: Evidence from Spain".

1303 *Cecilia Dassatti Camors:* "Macroprudential and monetary policy: Loan-level evidence from reserve requirements".

1304 *Juan Carvajal:* "Choosing to invest in human capital through adult education".

1401 *Haritz Garro:* "Desire to win and public information in majority rule elections".

1402 *Álvaro Martín Herrero:* "Credit and liquidity risk in sovereign bonds".

1403 *José Carreño:* "Housing bubbles, doubts and learning".

1404 *Ester Núñez de Miguel:* "Excuse me, do you speak English? An international evaluation".

1501 *Alexander Heinemann:* "Efficient estimation of factor models with time and cross-sectional dependence".

1502 *Ilya Morozov:* "The role of commitment in the U.S. movie industry".

1601 *Stefano Pica:* "Food versus non-food consumption insurance in Uganda".

1602 *Aliaksandr Zaretski:* "Optimal macroprudential rules".

1603 *Carlos Victoria Lanzón:* "Is Spain fiscally partisan? Evidence from local elections".

1701 *Miguel Ruiz:* "Bilingual education: Experience from Madrid".