Improving students' reading habits and solving their early performance cost exposure: evidence from a bilingual high school program in the Region of Madrid*

José Montalbán Castilla†
Paris School of Economics

May 2016

Abstract

Existing literature has highlighted the potential high returns of speaking a second language in many aspects of the labor market, incentivizing policy-makers to contribute to the the students' language skills improvement with public bilingual programs. Using a unique survey of 9th graders students in Spain, this paper provides evidence about the effects of a public English-Spanish bilingual program on students' performance and other students' outcomes. Under this program, students study different subjects in English in addition to English as a foreign language, with two differentiated programs starting in secondary education that differ in the intensity of the instruction time in English. I find no statistically significant impact on students' cognitive skills neither on subjects taught in English nor in Spanish. In contrast, I do find that the program has a strong positive impact on students' reading enjoyment time and reading habits. I also find relatively small negative effects on several high school students' perceptions mainly for children whose parents have less than college education, and no effects on students' opinion of teachers.

Keywords: Bilingual education, program evaluation, teaching in English.
JEL codes: H40, I21, I28.

*I am grateful to Ismael Sanz Labrador, Luis Pirez Jiménez, María Ángeles Díez Santos, Vicente Alcañiz Miñano, Francisco Javier García Crespo, David Cervera, Juan Manuel Escuderos, Luis Fernández Martín, Sebastián Huerta and Guillermo Gil Escudero for obtaining the data, their useful comments, guidance and help. Moreover, I would like to thank Brindusa Anghel, Julien Grenet, Clara Martínez Toledano and Alessandro Tondini for their devoted and useful help. Finally, I am also grateful to Fundación Ramón Areces for its financial support.

†Paris School of Economics, Boulevard Jourdan, 48, Paris 75014, France. *Email:* j.montalbancastilla@gmail.com

1 Introduction

Proficiency on at least a second language of wider world communication (especially English, although Spanish or French are also important) is required for future students to be able to meet the globalized labor market needs. The development of global interconnections in the last decades made crucial the proficiency of speaking in different languages for having a higher salary, and to increase the probability of finding a job, since most of the world companies have included language skills as one of the core requirements for hiring their workers. Ginsburgh and Prieto-Rodriguez (2011) found large estimates on the effects of foreign language knowledge on wages in Mincerian regressions, ranging the effects from 11% in Austria to 30% in Spain for English language knowledge. Sainz and Zoido (2005) found a wage premium effect of about 2%-3% for college graduates who can speak a second language in the US [1]. Furthermore, the cognitive psychology literature finds that bilingual education, especially early in life, has a positive impact on cognitive development, mostly working through improved executive functions and their corresponding brain structures (Costa and Sebastián-Gallés, 2014). However, the evidence on literacy rates of bilinguals is mixed with some studies pointing towards a negative effect on literacy skills measured right after exposure to bilingual education (Bialystock, 2007).

These potential high returns have incentivated policy-makers to contribute in the students' language skills improvement. However, implementing such a policy is not trivial. Increasing the time spent in foreign languages using the current teaching methods would disproportionately rise the share in the available curriculum time. Moreover, switching the entire academic curriculum to a different language may cause non-negligible costs. For instance, it can dampen students' knowledge of their mother tongue language and damage their understanding degree of the subjects. The current theory on second language acquisition (see Brown, 2000; Ellis, 2003; Lightbown and Spada, 1999; Mitchell and Myles, 1998; Richards and Rodgers, 2001) claims that meaningful and authentic communication that is functional and relevant to the students'

[1] Fidrmuc and Fidrmuc (2009) also showed the important role that English plays for foreign trade.

needs, and not contrived, tightly controlled or repetitive, is important for second language acquisition. Then, it is a prime order priority for policy-makers to find efficient paths to foster language skills.

The goal of this paper is to test whether bilingual educational immersion programs [2] have a cost in terms of secondary students' performance and other non-cognitive characteristics (such as students' reading enjoying time, reading habits, high school climate perception, and students' opinion of teachers). This potential loss may be caused by a worsening in students' knowledge of their mother tongue language and/or by weakening students' understanding degree of the subjects. The analysis is conducted in the context of Madrid Region's public bilingual program. Several Spanish administrations have offered bilingual education in schools across the country. However, the most ambitious program is developed by the autonomous region of Madrid [3]. This project started in the academic course of 2004/2005 in primary education for 26 schools, and in the academic year of 2015/2016 it had 353 public schools and 110 public high schools. In this program between 30% and 50% of the instruction is taught in English for primary education, and at least one third for secondary education. Mathematics and Spanish Language are always taught in Spanish, while the rest of subjects can be in English (usually the science curriculum). With respect to the existing literature, the contribution of the paper is twofold: (i) First, I develop the study using a unique pilot survey collected by Madrid Region in the course 2013/2014. (ii) Second, the paper analyzes bilingual program impacts on students' performance and other characteristics not related to cognitive skills.

To test the effects of the program I use the Programme for International Student Assessment (PISA) standardized students' results elaborated by the Organization for Economic Co-operation and Development (OECD). The exam evaluates language, mathematics and science skills for 15 years old students every three years. I use Madrid

[2]The term bilingual education in this case refers to 'immersion' , in which a foreign language (language which is not the majority mother tongue of the society) is the medium of instruction.

[3]There exists also an agreement between the Ministry of Education and the British Council that selects 80 schools all over Spain. Moreover, bilingual programs in several Spanish regions are developed, but not with the intensity of Madrid. For instance, in Andalusia there is a bilingual program but with very small percentage of instruction in English (around 20%) .

regional data of PISA 2009 (before the treatment) for 19 high schools where two of them implemented the bilingual program on the next academic year (in 2010/2011). I merge them with a pilot survey (with the same structure of PISA) developed by the Regional Government of Madrid in the first quarter in the school year of 2013/2014 and I checked the results of this cohort.

The paper faces a triple potential selection problem: high schools applying to the program, students choosing high schools and selection of students' allocation within the two bilingual programs (*Bilingual Section* and *Bilingual Program*). I propose several solutions to facing them. For high school selection, I compare the same high schools before and after the program was implemented. Moreover, I attempt to solve students potential selection bias by controlling for different key observable students' characteristics, and by eliminating the selection bias that does not change over time. The main empirical strategy is a Difference-in-Difference (Diff-in-Diff) approach with controls, comparing the exam performance of students in the treated high schools before and after the program was implemented with the group of non-bilingual high schools before and after the treatment. I do not find statistically significant effects of the program on Reading, Mathematics and Sciences performance either by high school or students'. Heterogeneous effects are investigated for students whose parents' report a university degree and the rest, not finding significant results either making separate regressions by group or developing a Triple Diff-in-Diff strategy in order to maximize the sample size. The same is found for high school dispersion of performance. In contrast, I do find strong positive and significant effects on students' reading enjoying time of 16.5 pp (137 percent of the baseline mean), driven by children of parents with a college education. I do also find evidence of positive effects of the bilingual program on beneficial reading habits and no effect on negative reading habits, increasing students' opinion of reading as a favorite hobby in 17.5 pp (87.5 percent of the baseline mean) and enjoy the bookstore in 17 pp (212 percent of the baseline mean). The impacts are higher for students' whose parents have college education, and the rest of students report also significant and positive impacts on the opinion of enjoying exchanging books with partners. Moreover,

I find relatively small negative and significant results on students' perception that high school classes improve students' confidence on 15.9 pp (19 percent of the baseline mean). This result is explained by students whose parents do not report university education who display an additional negative effect of 18.4 pp (18.4 percent of the baseline mean) in the opinion that high schools are not useful for finding a job. Finally, I do not find significant results of the program impact on students' opinion of teachers.

There is little evidence regarding the effects of a bilingual education in English for countries whose official language is not English. Anghel et al (2015) evaluate the bilingual program in Madrid for primary schools finding for the first treated cohort a not significant effect different from zero for either Mathematics and Spanish language. Although for General Knowledge, the bilingual program has a negative and significant effect on exam results for children of parents without a college education. However, the effects for the schools that started to participate a year after the beginning of the program are also significant and negative, but smaller in size [4]. In addition, Admiraal et al (2006) study the impact of a bilingual program of English for secondary education in the Netherlands, finding no effects for receptive word knowledge and no negative effects with respect to the results of their school leaving exams at the end of secondary education for Dutch and subject matters taught in English.

The remainder of this paper is structured as follows. The next section is devoted to explain the Madrid Region high school bilingual program. Section 3 presents the data. Section 4 shows the descriptive statistics. Section 5 displays the selection problems and the empirical strategy. Section 6 analyzes the impact of the bilingual program on students' performance, high school dispersion, and other students' outcomes. Section 7 concludes.

[4]Hoya (2015) find when analyzing the same program that "switching to bilingual education had no significant detrimental effect on students' average performance in the courses that continued to be taught in Spanish in the bilingual program, but there is suggestive evidence that learning English might come at the cost of some technical knowledge and skills in courses that begin taught in English", but the analysis of Anghel et al (2015) uses more powerful techniques for controlling the selection problems.

2 Program description

The Spanish educational system is formed by six years of primary school, four years of compulsory secondary education (*E.S.O*), and two years of non-compulsory education divided into preparation for college (*Bachillerato*) and vocational training (*Ciclos Formativos*).

The bilingual project in the Region of Madrid started in 2004/2005 as an experimental program for first grade students of 26 selected schools. Students from treated schools from cohorts starting just before this academic course did not participate in the program and continue with the same program as in the untreated schools. Successive cohorts of treated schools have also been treated. Besides, additional primary schools joined the program in the successive years, counting with 353 public bilingual schools in the course 2014/2015. Due to the high demand of the program, once the students from the 2004/2005 cohort reached secondary education in the course 2010/2011, a second phase started and 32 high schools joined the program. Nowadays, bilingual program is working in 110 public high schools of Madrid Region.

Each bilingual high school must organize their teaching in two modalities: bilingual program (*programa bilingue*) and bilingual section (*sección bilingue*).

1. *Bilingual Program:* it is based on the increment of two weekly hours of English from the general timetable, in such a way that there is always an English class per day. Moreover, it is possible to teach English in other subjects such as Physical Education, Art, Music or Citizenship Education, as well as the tutorials for secondary. The high schools can organize flexible groups depending on the English level of the students as well as extracurricular activities with the goal of promote the bilingual atmosphere.

2. *Bilingual Section:* this program has two sections. i) Teaching of English: teaching English the five school days of the week, an hour per day. ii) Teaching in English: students may study all the courses in English with the exception of Mathematics, Spanish language and Literature, and the second language they choose. In any

case, the same subject cannot be taught in both languages. Moreover, tutorials must be taught in English. In the first, second and third course of secondary, Social Sciences, Geography and History, and Sciences must be taught in English, as well as at least Biology and Geology (in the third course). In the last secondary course high schools must add an additional course in English. The English teaching and subjects taught in English must be at least one third of the weekly school time.

The student admission procedure is done by the regulation of Madrid Region. However, bilingual high schools may use a complementary admission criterion in order to guarantee the viability of the program. Meaning that as a general rule students are in Bilingual Program. However, high school's executive team determines the inclusion of each student to the Bilingual Section based on students' academic files, students' performance in an external test of English skills in the last course of primary education, CDI test of Madrid Region (Indispensable Knowledge and Skills), and any other criteria that determines new students' English knowledge. Furthermore, on the course 2013/2014, the admission procedure for Bilingual Section was modified [5]. Students' who came from a public bilingual school must pass the four categories of the external test for English skills in the last course of primary education. Students' who do not came from a public bilingual school must provide a certificate proving that they count with a B1 English level (Common European Language Framework), and in some extraordinary justified cases the direction of the high school can ask for making a level test.

The high school students' are evaluated at the end of the second and fourth course of secondary education by the Education Office of Madrid Region.

Teachers in bilingual high schools receive an special complement over their basic wage for the "special performance developed by the teachers, the extra dedication that results in a longer workday, due to the higher demands imposed by the activities of class preparation, processing and adaptation of materials into other languages, and

[5] B.O.C.M number 206. July 31th 2013 Order.

regular attendance at coordination meetings outside school hours" [6]. The complement vary from 41 to 215 additional euros per month depending on the number of hours taught in English and coordination tasks. It means 1.958 additional euros per year for a teacher who teaches more than 10 hours per week in English (for subjects different than English language); a teacher who teaches between 7-10 hours, 1.468; between 4-6 hours, 979; and between 1-3 hours, 489. The program also provides "conversation assistants" to high schools, who are usually college students from English speaking countries. Finally, in order to teach in English, teachers should be either specialists in English (proving at least a C1 level on the Common European Language Framework) or pass an exam that test their English skills. The exam contains two parts: written exam (reading, writing, listening, vocabulary and grammar) and an oral part.

[6]B.O.C.M number 278, November 14h 2006 Order.

3 Data

The *Programme for International Student Assessment (PISA)* international study is elaborated by the Organization for Economic Co-operation and Development (OECD) with the goal of evaluating language, mathematics and science skills for 15 years old students every three years. Moreover, PISA compiles data on students' family context, schools and educational system. Spain participated in the study in 2000 for the first time, and in 2009 the regional governments had the opportunity to increase the sample with the aim of making them comparable with the rest of countries and regions. Madrid Region raised the sample in this year. *PISA for Schools* is a complement to the national PISA evaluation run in 2009. This program shares the theoretical framework of evaluation with PISA, but the evaluations are different. It is a students' evaluation tool mainly directed to schools and schools associations. It contains descriptive information, and an analysis of skills and practical application of knowledge of 15 years old students on Mathematics, Reading and Sciences. Furthermore, it also provides information about school internal factors, socioeconomic characteristics of students, students' attitudes and interests. However, these data are comparable with PISA scales.

In addition to this dataset, I use Madrid regional data on PISA 2009 merged with a pilot study developed by the Regional Government of Madrid in the first quarter of the course 2013/2014. The design of this study is inspired in the theoretical evaluation framework of PISA for Schools and the results obtained are comparable.

The goal is to evaluate the same high schools before the implementation of the bilingual program (2010/2011 course for secondary education) and after (2013/2014).

4 Descriptive Statistics

Table 1 presents the differences in the students' observable variables between bilingual and non-bilingual high schools before the treatment, in 2009. The goal is to show that treatment and control groups are very similar before the treatment in observable characteristics. We do find that both groups report statistically significant differences in the proportion of female students, socioeconomic index (ESCS), proportion of parents who have an university degree, all parental occupations (business, professional and blue collar), home possessions index, and lessons students' self-declared time per week. A problem arises for the comparability of the two groups, meaning that a considerable amount of students' characteristics are different in treatment versus the control group, before the treatment.

Table 1: Descriptive Statistics. Differences in observable variables between treatment and control students in 2009.

	Mean (T)	**SD (T)**	**Mean (C)**	**SD (C)**	**Difference**	**P-value**
Female	.6	.494	.479	.500	-.121	.078
Age	15.88	.282	15.87	.282	-.009	.820
Indexes						
Home possessions	.867	.110	.818	.133	-.049	.0063
ESCS	.378	10.36	-.051	10.64	-.429	.0032
Famstruc	17.62	.43	18.44	.38	.082	.126
Family members at home						
Mother at home	.983	.130	.989	.309	.005	.902
Father at home	.821	.386	.890	.429	.069	.251
Brother at home	.62	.490	.560	.592	-.061	.501
Sister at home	.542	.503	.592	.589	.051	.564
Parental education						
University	.533	.503	.414	.493	-.119	.077
Higher secondary	.233	.426	.247	.432	.014	.81
Vocational training	.067	.251	.105	.307	.038	.353
Lower secondary	.017	.130	.0159	.125	-.0008	.961
Parental occupation						
Business	.4	.494	.279	.449	-.121	.052
Professional	.183	.391	.109	.312	-.074	.090
Blue collar	.133	.343	.315	.465	.181	.003
Lessons Time						
Language	50.83	2.09	52.95	4.89	2.12	.0009
Mathematics	50.67	2.83	53.01	4.96	2.34	.0003
Science	50.8	2.55	52.86	4.89	2.06	.0035
Number of weekly lessons						
Language	4.25	.756	4.27	.854	.014	.905
Mathematics	3.47	1.33	3.66	.986	.184	.193
Science	4.32	1.97	3.92	1.86	-.4	.150
Obs. Students	60	60	505	505		
Obs. High schools	2	2	17	17		

Table 2 shows the same student's variables means for the treatment and control groups in both years, and the raw Diff-in-Diff. Students' characteristics in treatment and

control groups seem to report a common trend in almost all the covariates. It means that some observable students' characteristics were different in control and treatment groups before the treatment, but most of these differences are constant over time, and the Diff-in-Diff analysis solves this problem. The variables which are not comparable between the two groups are ESCS, students with parents working on professional and blue collar occupations, and lessons time per week in Mathematics and Reading (differences that seems reasonable by the creation of the bilingual program). This implies that for these variables the selection bias did not follow a common trend, and the changes in these five variables along time capture the effect of the bilingual program plus other confounding changes. I attempt to control for these confounding factors by adding these variables in the regression, and variables which are related with the ESCS as parental education and occupation.

Table 2: Descriptive Statistics benchmark (standard errors in parenthesis)

	Year 2009		Year 2013			
Variables	**Bilingual**	**Non Bilingual**	**Bilingual**	**Non Bilingual**	**Diff-in-Diff**	**P-value**
Female	0.600	0.479	0.521	0.479	-0.079	0.326
	(0.494)	(0.500)	(0.501)	(0.500)	(0.080)	
Age	15.88	15.87	15.49	15.43	0.047	0.297
	(0.283)	(0.282)	(0.286)	(0.285)	(0.045)	
Indexes						
Home possessions	0.867	0.818	0.831	0.809	-0.027	0.215
	(0.110)	(0.133)	(0.138)	(0.137)	(0.022)	
Famstruc	1.763	1.845	1.820	1.829	0.073	0.242
	(0.429)	(0.384)	(0.385)	(0.392)	(0.063)	
ESCS	0.378	-0.0509	0.437	0.314	-0.306	0.045
	(-1.036)	(-1.064)	(0.833)	(0.904)	(0.152)	
Family members at home						
Mother at home	0.983	0.988	0.976	0.978	0.003	0.921
	(0.130)	(0.309)	(0.153)	(0.148)	(0.033)	
Father at home	0.821	0.890	0.844	0.834	0.080	0.211
	(0.386)	(0.430)	(0.364)	(0.373)	(0.064)	
Brother at home	0.622	0.560	0.509	0.517	-0.070	0.452
	(0.490)	(0.592)	(0.501)	(0.500)	(0.093)	
Sister at home	0.542	0.593	0.509	0.517	0.043	0.638
	(0.504)	(0.590)	(0.501)	(0.500)	(0.091)	
Parental education						
University	0.533	0.414	0.538	0.511	-0.092	0.248
	(0.503)	(0.493)	(0.500)	(0.500)	(0.080)	
Higher secondary	0.233	0.248	0.166	0.155	0.025	0.687
	(0.427)	(0.432)	(0.373)	(0.362)	(0.062)	
Vocational training	0.0667	0.105	0.148	0.117	0.069	0.177
	(0.252)	(0.307)	(0.356)	(0.322)	(0.051)	
Lower secondary	0.0167	0.0158	0.0533	0.0751	-0.023	0.529
	(0.129)	(0.125)	(0.225)	(0.264)	(0.036)	
Parental occupation						
Business	0.400	0.279	0.485	0.412	-0.048	0.537
	(0.494)	(0.449)	(0.501)	(0.492)	(0.077)	
Professional	0.183	0.109	0.154	0.187	-0.107	0.067
	(0.390)	(0.312)	(0.362)	(0.390)	(0.059)	
Blue collar	0.133	0.315	0.219	0.261	0.140	0.048
	(0.343)	(0.465)	(0.415)	(0.439)	(0.071)	
Lessons time						
Language	50.83	52.95	60	60	2.119	0.000
	(2.092)	(4.897)	(0)	(0)	(0.417)	
Mathematics	50.67	53.01	60	60	2.342	0.000
	(2.832)	(4.959)	(0)	(0)	(0.427)	
Science	50.80	52.86	57.84	58.32	1.586	0.295
	(2.548)	(4.894)	(11.20)	(9.902)	(1.513)	
Number of weekly lessons						
Language	4.254	4.268	4.156	4.192	-0.023	0.809
	(0.756)	(0.854)	(0.364)	(0.413)	(0.094)	
Mathematics	3.475	3.659	3.760	3.774	0.170	0.176
	(1.331)	(0.987)	(0.623)	(0.641)	(0.126)	
Science	4.320	3.919	4.048	4.124	-0.476	0.059
	(1.974)	(1.859)	(1.388)	(1.250)	(0.252)	
Obs. Students	60	505	169	1,039		
Obs. High schools	2	17	2	17		

5 Empirical strategy

Ideally, students' performance in the standardized exam would be represented by regressing the score of a student in a given high school on a dummy indicating whether the high school program is bilingual (BIL_j). For a given year, the score in the test for student i in high school j, y_{ij}, is determined by:

$$y_{ij} = \alpha + \beta BIL_j + \epsilon_{ij} \tag{1}$$

where ϵ_{ij} is a random shock. The parameter of interest would be β, which captures the average treatment effect of the bilingual program on students' performance. If the assignment of the program across high schools and the allocation of the students on the different high schools would be random or as good as random, this regression would identify the causal effect of the bilingual program. However, both of these premises are not true in this case. This model suffers from several potential endogeneity bias sources by three self-selection problems (the two first ones are also faced by Anghel et al (2015)):

1. *High schools are not randomly selected to implement the bilingual program*. The program was implemented only in some of the high schools that applied for it. The extra amount of work required by the program could lead to interpret the application to the program as teachers and principal high quality signal.

2. *Students are not randomly assigned to high schools*. The parents choose the high school. If there is no excess of demand for the high school they have chosen, they are admitted. If there is excess of demand, the admission is based on proximity to high school and family income, both of which are not random and correlated with high school outcomes.

3. *Allocation of students between the two programs is not random*. Bilingual high schools may use a complementary admission criterion based on students' academic files, students' performance in an external test of English skills, CDI test of Madrid Region, and any other criteria that determines new students' English knowledge.

Then, the model suffers from a potential selection bias problem given the fact that all of this criteria are correlated with student's performance.

Furthermore, a potential bias problem arises by the fact that students can be exposed to different intensities of treatment, but the researcher cannot observe it. Students in bilingual high schools are divided in Bilingual Program and Bilingual Section, both of them different in the intensity of English instruction and the subjects they provide (as I indicate in Section 2). In the course 2013-2014, the high-school percentage average of nine-graders students in Bilingual Program classes was 71.5% in the Region of Madrid (over the total bilingual high-schools). The rest were located in Bilingual Section classes.

In an attempt to control for the selection problems mentioned above, I use a Difference in Differences estimation (Diff-in-Diff) as the main identification strategy to estimate the effect of the bilingual program on students' outcomes. This strategy solves the self-selection of high schools into the program, since it compares the same high schools before and after the program implementation. I am going to estimate the following regression by OLS:

$$y_j = \alpha + \beta BIL_j + \gamma year2013 + \delta(year2013 * BIL_j) + \epsilon_j \quad (2)$$

where $year2013$ is a dummy for the academic course 2013/2014. The parameter of interest is δ which captures the average treatment effect of the bilingual program on high schools' scores. However, this model does not attempt to control for the individual differences in observable characteristics over time. I cannot assure that students self-selection does not change along time.

To solve the second source of endogeneity, it is necessary that students' movements in bilingual high schools after the program were implemented, were the same as in the absence of the program. Nonetheless, since the program was at the first stage implemented in primary education in the course 2004/2005, students' families had a considerable amount of time to sort students in bilingual high schools. Then, I cannot assure that Diff-in-Diff solves completely the second selection problem, but model (3)

specification includes observable students' characteristics, in an attempt to control for students' characteristics changes between treated and untreated high schools. Moreover, I do not have enough information to figure out the potential source of bias since the researcher cannot distinguish which of the two programs is being taught for each student in bilingual high schools. However, we suspect that likely a large fraction of bilingual students were in Bilingual Program instead of Bilingual Section (based on the general high proportion of students' in that Program).

Given the main identification strategy, I am going to estimate the following regression by OLS:

$$y_{ij} = \alpha + \beta BIL_{ij} + \gamma year2013 + \delta(year2013 * BIL_{ij}) + \eta x_{ij} + \epsilon_{ij} \quad (3)$$

where x_{ij} is a vector of students' observable characteristics. The parameter of interest is δ which captures the average treatment effect of the bilingual program on students' scores. The main identification assumption is that any change in students' performance between students in the control group and students in the treatment group is due to the bilingual program.

Finally, it is interesting to develop an heterogeneous effect analysis by parental education. It is a key variable given the fact that it is a predetermined variable, and it is a crucial affecting students' performance (much more than parental occupation or parental wealth) [7]. I run separately equation (3) for students whose parents have a college degree, and the rest. However, given the statistical power problems that this analysis reports by dividing the sample in two, I also test a Difference-in-Difference-in-Difference analysis (Triple Diff-in-Diff). The goal of this model is to analyze whether there exists differences on the effect of the bilingual program by two groups without limiting the sample size. I estimate the following regression by OLS:

[7] I do not elaborate this analysis with more groups because of sample size restrictions.

$$y_{ij} = \alpha + \beta BIL_{ij} + \gamma year2013 + \delta(year2013 * BIL_{ij}) + \zeta univ_{ij} + \lambda(univ_{ij} * BIL_{ij}) + \sigma(year2013 * univ_{ij}) + \phi(year2013 * BIL_{ij} * univ_{ij}) + \eta x_{ij} + \epsilon_{ij} \quad (4)$$

where $univ_{ij}$ is a dummy equal to one if the student i of high-school j has a parent with university degree, and zero otherwise. The parameter of interest is ϕ which captures the difference between the average treatment effect of the bilingual program between students' with college parental education and the rest.

6 Results

6.1 Impact on cognitive skills

Bertrand et al (2004) addressed the importance of potential serial correlation in the Diff-in-Diff estimation which derives in inconsistent standard errors. For the cases where the clusters are few, they propose to ignore the time series information, implementing a "correction that collapses the time series information into a "pre" and "post" period". As I count with a few number of clusters (19), I start by investigating the effect of bilingual program on high schools with no cluster corrections. Table 3 shows the Diff-in-Diff results by high schools. I do not find statistically significant results of a difference in performance on any of the three subjects by the bilingual program: Spanish language, Maths and Sciences. Furthermore, the sample is restricted to the same number of treatment and control high schools (2). In order to improve the comparability of both groups, I choose the two control high schools which are the most similar to the treatment in students' observable variables, before the implementation of the program. The coefficients are much more higher than before (with very similar standard errors) and the R-squared is bigger, but still they remain non significant.

Table 3: Diff-in-Diff results by high schools

Variables	**Reading**		**Mathematics**		**Sciences**	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	511.368***	516.474***	502.624***	501.305***	513.156***	521.267***
	(9.867)	(22.214)	(10.265)	(14.850)	(9.370)	(20.933)
Year 2013	5.540	14.490	-4.432	18.254	-7.132	-3.045
	(13.954)	(31.415)	(14.517)	(21.002)	(13.251)	(29.604)
Bilingual High-school	-1.908	-7.014	3.592	4.910	3.206	-4.905
	(30.413)	(31.415)	(31.640)	(21.002)	(28.880)	(29.604)
Bilingual High-school	-8.645	-17.595	0.102	-22.584	0.756	-3.331
in 2013	(43.010)	(44.428)	(44.745)	(29.701)	(40.842)	(41.867)
Controls	No	No	No	No	No	No
Obs. High Schools	19	4	19	4	19	4
R-squared	0.007	0.154	0.004	0.197	0.010	0.037

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The analysis developed above does not attempt to control for the students' individual differences in observable characteristics. As I face two potential selection problems by using model (2), I cannot assure that students self-selection does not change along time. As I show in the descriptive statistics, the treated students have different characteristics

than untreated, and those characteristics affect positively the students' performance. I attempt to control the self-selection of students by using the model (3), a Diff-in-Diff of students' results with covariates.

The dependent variable of this Diff-in-Diff analysis is computed with plausible values using Rubin's (1987) combination methods. The statistic is calculated for each plausible value and then averaged. Its sampling variance is based on the variation between the statistic calculated for each plausible value. This procedure is specific for several databases (including PISA data). For the replication variance estimator, I use the Balanced Repeated Replicates (BRR) replication method for computing the standard errors, recommended by Kreuter et al (2007) for PISA data.

Table 4 presents the estimates of the model (3). The parameter δ estimates the average treatment effect of the bilingual program. I do not find statistically significant results of a difference in performance given by the bilingual program for any of the three examination subjects: Spanish language, Maths and Sciences. Students' performance in bilingual high school do not seem to be damaged or to have benefited by the program in comparison with students in the regular Spanish program.

Table 4: Diff-in-Diff results for students with and without covariates. All students

Variables	Reading		Mathematics		Sciences	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	522.833***	431.445***	517.078***	617.951***	527.483***	422.688**
	(11.194)	(164.483)	(11.525)	(196.696)	(11.056)	(190.652)
Year 2013	-0.270	-3.003	-9.509	-20.297**	-16.624	-15.353
	(11.662)	(9.519)	(11.603)	(9.007)	(11.039)	(9.631)
Billingual high	-1.890	-10.949	13.527	-6.406	13.250	-6.178
school	(28.661)	(10.874)	(31.140)	(12.290)	(34.256)	(14.250)
Bilingual high	-13.811	-9.950	-18.996	-6.182	-14.127	-1.226
school in 2013	(29.357)	(12.406)	(32.105)	(13.846)	(34.736)	(15.436)
Female		-32.356***		6.079		0.292
		(8.563)		(8.173)		(8.689)
Student's age		2.277		-13.178		1.200
		(8.808)		(10.322)		(12.090)
Home possessions		95.046***		108.503***		125.718***
		(31.502)		(40.941)		(35.050)
Mother at home		24.401		45.055		32.447
		(19.039)		(27.516)		(22.702)
Father at home		1.109		3.240		-4.185
		(8.843)		(9.472)		(9.725)
Brother at home		-9.076		-7.648		-15.544**
		(6.991)		(8.948)		(7.327)
Sister at home		-15.059**		-29.281***		-14.770*
		(6.891)		(8.359)		(8.312)
University		42.020***		41.504***		29.197*
		(14.955)		(14.100)		(15.514)
Higher secondary		5.996		11.249		1.747
		(10.028)		(12.128)		(11.424)
Vocational training		5.600		4.833		-2.980
		(11.651)		(14.949)		(13.942)
Lower secondary		-19.750		-27.823		-13.350
		(25.713)		(22.881)		(20.978)
Business		14.634		25.415*		12.253
		(13.257)		(14.387)		(12.293)
Professional		-7.105		-7.226		-10.655
		(11.736)		(10.630)		(11.807)
Blue collar		-20.333		-16.170		-33.539**
		(14.336)		(14.556)		(14.409)
Time of Reading		-1.765		-1.784		-2.096
		(2.686)		(2.988)		(3.003)
Time of Math		2.740		2.951		2.750
		(2.806)		(3.120)		(3.226)
Time of Science		-1.198*		-0.958		-1.141**
		(0.685)		(0.646)		(0.519)
Number of Reading		-12.703		-21.167		-10.755
		(13.112)		(13.808)		(11.676)
Number of Math		-6.120		-3.482		-5.820
		(7.592)		(8.389)		(6.895)
Number of Science		12.582***		13.608***		13.739***
		(2.656)		(3.038)		(2.674)
Obs. students	1,510	1,510	1,510	1,510	1,510	1,510
Obs. High school	19	19	19	19	19	19
Average R-squared	.0006	.351	.003	.351	.006	.352

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5 reports the results of the estimated average treatment effect by two groups of parental education: students whose parents obtained an university degree and the rest. I still do not find statistically significant results for any of the subjects. I do find a relatively similar effect in performance between students of both groups. The only discipline where both students' groups perform clearly differently is Mathematics,

where students whose parents report a university degree seem negatively affected by the program, and the rest positively stimulated. However, the coefficients are not statistically significant.

Table 5: Separate Diff-in-Diff regressions for observable groups of students: estimated treatment effect by group.

	Reading		**Mathematics**		**Science**	
University	-18.570	-3.871	-32.764	-15.154	-17.419	5.299
	(23.767)	(13.246)	(27.427)	(16.583)	(26.645)	(15.548)
Controls	No	Yes	No	Yes	No	Yes
Obs. students	760	760	760	760	760	760
Average R-squared	.011	.269	.012	.305	.027	.323
					.	
Non-University	10.035	-3.550	16.310	15.227	8.412	2.999
	(17.133)	(16.023)	(14.883)	(15.332)	(19.583)	(15.616)
Controls	No	Yes	No	Yes	No	Yes
Obs. students	750	750	750	750	750	750
Obs. High schools	19	19	19	19	19	19
Average R-squared	.002	.326	.001	.266	.001	.296

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6 shows the results of the model (4) described in Section 5. The last row displays the estimated average treatment effect differences between students whose parents has university degree and the rest. I observe non significant different effects between the two groups. Then, Table 5 and 6 do not report significant evidence to support the idea that the bilingual program provide an additional effect for students whose parents report university studies.

Table 6: Triple Diff-in-Diff results for students with and without covariates. All students

Variables	**Reading**		**Mathematics**		**Sciences**	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	494.266***	436.530***	487.282***	629.669***	500.142***	426.520**
	(11.208)	(162.471)	(9.763)	(193.760)	(9.988)	(179.681)
Bilingual High-school	-13.897	-12.692	-3.378	-14.507	-3.324	-14.857
	(14.917)	(12.566)	(11.920)	(10.948)	(20.997)	(16.040)
University	66.312***	40.514***	69.165***	36.792***	63.464***	30.920**
	(18.156)	(14.503)	(19.463)	(14.060)	(17.461)	(14.227)
Year 2013	2.920	5.106	-10.973	-16.126*	-8.860	-2.901
	(12.509)	(12.829)	(10.178)	(9.367)	(10.057)	(9.050)
University in	4.085	2.617	11.862	14.359	12.728	15.006
bilingual High-school	(24.617)	(18.755)	(26.089)	(21.421)	(24.721)	(20.313)
Bilingual High-school	10.035	-1.566	16.310	15.081	8.412	6.478
in 2013	(17.133)	(16.226)	(14.883)	(15.008)	(19.583)	(15.328)
University	-16.946	-18.055	-8.355	-12.104	-25.384	-26.945*
in 2013	(19.616)	(16.963)	(19.830)	(16.965)	(17.494)	(14.347)
University in bilingual	-28.605	-12.479	-49.074	-35.902	-25.831	-11.327
high-school in 2013	(29.197)	(24.363)	(30.871)	(27.923)	(27.044)	(21.886)
Controls	No	Yes	No	Yes	No	Yes
Obs. students	1,510	1,510	1,510	1,510	1,510	1,510
Average R-squared	.146	.350	.149	.348	.142	.354

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6.2 Impact on Students' cognitive skills dispersion

I do not find any statistically significant result on students' performance, either by high school and students' individually. However, the bilingual program may affect the students' performance dispersion at high school level. The result of a non-statistically significant effect by high-school on students' cognitive skills, do not imply that the program had not affected the high school dispersion of performance, suggesting a higher or lower heterogeneity between high schools students'. Table 7 displays the results of the bilingual program impact on high school average standard deviation of students' performance. I do not find statistically significant results on cognitive skills dispersion by high school, suggesting that the distribution of students' performance (at the high school level) did not changed by the implementation of the program.

Table 7: Diff-in-Diff results for dispersion analysis: estimated treatment effect by group

	All high schools in the sample			**Most comparable high schools**		
	Reading	**Math**	**Sciences**	**Reading**	**Math**	**Sciences**
Constant	4.175***	4.752***	4.520***	5.554**	2.676	5.520**
	(0.493)	(0.433)	(0.333)	(1.530)	(1.842)	(1.633)
Year 2013	0.074	-0.113	-0.755	-1.383	0.881	-1.475
	(0.697)	(0.613)	(0.471)	(2.163)	(2.605)	(2.309)
Bilingual High-school	0.011	0.514	-0.427	-0.288	1.509	-1.427
	(1.519)	(1.336)	(1.026)	(2.163)	(2.605)	(2.309)
Bilingual High-school	0.740	-0.469	1.263	0.801	-0.067	1.983
in 2013	(2.148)	(1.889)	(1.451)	(3.059)	(3.684)	(3.266)
Obs. High school	19	19	19	4	4	4
R-squared	0.009	0.007	0.076	0.108	0.176	0.117

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6.3 Impact beyond cognitive skills

Madrid Region bilingual program states a change in the teaching structure that affects much more educational aspects than language of teaching. The plan involves teachers incentive policy (as described in Section 2) by increasing the wage compensation, changes the distribution of the different subjects over the total instruction time, include an external test of English skills, among other changes. It is crucial to evaluate other aspects beyond cognitive performance in order to have a clear picture of the different effects that the program may cause. The database allows to develop a unique contribution to the literature: test the effect of the program on other variables different than students' cognitive skills. I evaluate the program effects on students' reading enjoying time, reading attitudes, high school students' perception, and students' opinion of teachers .

In the course 2010/2011, students' who finished the primary education in the public bilingual schools of Madrid Region advanced to the public bilingual high-schools. A legislative order determined the curricula of *Advanced English* for the Bilingual Section of bilingual high schools [8]. This curricula commands the contents of the English teaching which be about English Literature and Language. It suggest that "the Order propose a new methodological approach, that implies teacher developing the contents of this subject through texts of literary nature. This innovation allows the correct development of the writing and oral skills, and the incorporation of the cultural and literary elements link to English language in a diary basis application"[9]. There are three main axis in the program: i) linguistic skills; ii) constitutive elements of the linguistic system and its relations; iii) cultural and social dimension of English language. Within the last one the two blocks are: sociocultural aspects and literary education. Then, it is an attempt to promote students' knowledge of cultural aspects of the English-speaking countries by focusing on the study of the literature production of representative writers of those countries. The students would "contact with a selection of literary texts with the goal of learning a variety of linguistic elements applied to their own writing skills, foment

[8] B.O.C.M number 160, July 6th of 2010. B.O.C.M number 160, July 8th of 2011. B.O.C.M number 214, September 7th of 2012.

[9] Order 2154/2010, April 20th Madrid Education Department.

their aesthetic sense, and contribute to their intellectual growing. Students would be familiarized with many styles through the different literary work that portrays cultural elements and historic contexts that shows the diversity of the English-speaking countries"[10]. It is clear that this change in the philosophy of teaching in bilingual high schools may affect the students' reading habits, their perception of English-speaking countries, their desire to know other cultures and societies, their self-confidence to travel abroad, among others. In the following analysis I focus on evaluating the effects on students' reading enjoying time, reading habits, students' high school perceptions, and students' opinion of teachers.

Table 8 displays the results of the regressions on reading enjoying time. In the questionnaire a question asks about how much time is devoted by the students' to read as hobby [11], with five possible answers: 1) don't read as hobby; 2) less than 30 minutes per day; 3) between 31 and 59 minutes per day; 4) between 1 and 2 hours per day; 5) more than 2 hours per day. I consider that more than 30 minutes per day is a reasonable indicator for a high reading enjoyment time at their age, and construct a dummy variable which is equal to one if the student's answer is 3,4 or 5, and zero otherwise [12]. I do find positive and significant results of the bilingual program on students' reading enjoyment time. Overall, the effect of the program is 16.5 pp (137 percent of the baseline mean). However, the effect is higher for students whose parents certify a university degree, reporting an effect of 26.6 pp (186 percent of the baseline mean), being non-significant for the rest of the students with parental lower education. I do not observe a significant effect for the extreme dummy's of "don't read" and "read more than 2 hours per day", but the coefficients seem to indicate that students' improve their reading enjoying time (although they are not statistically significant)

[10]Order 2154/2010, April 20th Education Department of Madrid.

[11]Question 09-ST23 in PISA 2009 and AP23 in 2013 data.

[12]The standard errors are corrected for heteroskedasticity because the dependent variable is heteroskedastic by definition.

Table 8: Diff-in-Diff results for reading enjoy time: estimated treatment effect by group

	Dummy total		**Dummy don't read**		**Dummy more than 2 daily hours**	
Total	0.240***	0.165*	-0.112	-0.053	0.034*	0.022
	(0.081)	(0.088)	(0.112)	(0.115)	(0.019)	(0.022)
Controls	No	Yes	No	Yes	No	Yes
Baseline mean	.12	.12	.48	.48	0	0
Obs. Students	1,510	1,510	1,510	1,510	1,510	1,510
Adjusted R-squared	0.00733	0.0611	0.00290	0.0590	-0.00123	0.0116
University	0.260**	0.266**	-0.126	-0.103	0.033	0.024
	(0.117)	(0.126)	(0.149)	(0.152)	(0.025)	(0.029)
Controls	No	Yes	No	Yes	No	Yes
Baseline mean	.143	.143	.428	.428	0	0
Obs. Students	760	760	760	760	760	760
Adjusted R-squared	0.00990	0.0374	0.00866	0.0390	0.00102	0.0207
Non-University	0.238**	0.134	-0.130	-0.058	0.043	0.019
	(0.110)	(0.107)	(0.167)	(0.166)	(0.032)	(0.038)
Controls	No	Yes	No	Yes	No	Yes
Baseline mean	.09	.09	.54	.54	0	0
Obs. Students	750	750	750	750	750	750
Adjusted R-squared	0.00295	0.0891	-0.00169	0.0665	-3.01e-05	0.00737
Triple Diff-in-Diff	0.022	0.091	0.004	-0.025	-0.010	0.000
	(0.161)	(0.173)	(0.224)	(0.227)	(0.041)	(0.044)
Controls	No	Yes	No	Yes	No	Yes
Obs. Students	1,510	1,510	1,510	1,510	1,510	1,510
Adjusted R-squared	0.00797	0.0536	0.00644	0.0523	0.000415	0.0115

Note:The coefficient reported in this table is the coefficient which captures the average treatment effect of the bilingual program on the dependent variable.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9 shows the results of the regressions on reading habits. In the questionnaire a question ask about several reading habits [13], with four possible answers from the question statement: 1) totally disagree; 2) disagree; 3) agree; 4) totally agree. I do divide the different reading habits on two groups: positive and negative attitudes. I consider as positive or beneficial the following sentences: reading is my favorite hobby, I like to talk about books, I am happy if I receive it as a present, I enjoy in the bookstore, I like to express my opinions about books, I exchange books with my classmates. In contrast, I consider as negative the following statements: I read only if I have to, its

[13]Question 09-ST24 in PISA 2009 and AP24 in 2013 data.

hard to finish a book, reading it's a waste of time, I read just because I need information, and I cannot sit still. I analyze these questions by creating a dummy variable which is equal to one if student's answer is 3 or 4 (they "agree" or "totally agree"), and zero otherwise. I do find evidence of strong positive effects of the bilingual program on students' reading habits. The program boosts students' opinion of reading as a favorite hobby in 17.5 pp (87.5 percent of the baseline mean) and enjoy the bookstore in 17 pp (212 percent of the baseline mean). Students whose parents report a university degree displays higher estimates of the two indicators mentioned before, 25.6 pp (88 percent of the baseline mean) and 28.3 pp (202 percent of the baseline mean) respectively, and incorporate a positive result on students' happiness when receiving a book in 31 pp (442 percent of the baseline mean). Furthermore, students whose parents do not have college education coincide with the rest in their positive impact of 12.7 pp on enjoying a bookstore, and incorporate an additional impact of 22.9 pp (127 percent of the baseline mean) on enjoying exchanging books with partners. In contrast, I find no significant effects on negative reading attitudes.

Table 9: Diff-in-Diff results for reading attitudes: estimated treatment effect by group

	Positives					
	Favourite hobbies	**Talk about books**	**Happy as present**	**Enjoy bookstore**	**Express opinions**	**Exchange**
Total	0.175*	-0.107	0.102	0.170**	0.023	0.135
	(0.099)	(0.116)	(0.094)	(0.072)	(0.110)	(0.106)
Baseline mean	.2	.52	.16	.08	.68	.28
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs. Students	1,510	1,510	1,510	1,510	1,510	1,510
Adjusted R-squared	0.0668	0.0663	0.0579	0.0760	0.0446	0.0855
University	0.256*	-0.093	0.310***	0.283**	0.026	0.139
	(0.147)	(0.147)	(0.108)	(0.119)	(0.146)	(0.153)
Baseline mean	.29	.64	.07	.14	.71	.35
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs. Students	760	760	760	760	760	760
Adjusted R-squared	0.0337	0.0405	0.0318	0.0651	0.0518	0.0804
Non-University	0.158	-0.052	-0.111	0.127*	0.075	0.229*
	(0.126)	(0.171)	(0.153)	(0.074)	(0.168)	(0.132)
Baseline mean	.09	.36	.27	0	.64	.18
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Obs. Students	750	750	750	750	750	750
Adjusted R-squared	0.0856	0.0773	0.0880	0.0859	0.0334	0.103

	Negatives				
	Only if I have to	**Hard to finish**	**Waste of time**	**Need information**	**Cannot l sit still**
Total	-0.014	0.031	-0.078	-0.000	0.026
	(0.114)	(0.114)	(0.099)	(0.111)	(0.083)
Baseline mean	.48	.36	.28	.4	.16
Controls	Yes	Yes	Yes	Yes	Yes
Obs. Students	1,510	1,510	1,510	1,510	1,510
Adjusted R-squared	0.0710	0.0134	0.0444	0.0722	0.0338
University	-0.144	-0.035	-0.024	-0.078	0.057
	(0.143)	(0.144)	(0.120)	(0.136)	(0.085)
Baseline mean	.43	.36	.21	.36	.07
Controls	Yes	Yes	Yes	Yes	Yes
Obs. Students	760	760	760	760	760
Adjusted R-squared	0.0442	0.000150	0.0211	0.0618	0.0232
Non-University	0.037	0.102	-0.153	0.043	-0.042
	(0.160)	(0.168)	(0.151)	(0.180)	(0.149)
Baseline mean	.54	.36	.36	.45	.27
Controls	Yes	Yes	Yes	Yes	Yes
Obs. Students	750	750	750	750	750
Adjusted R-squared	0.0725	0.00669	0.0676	0.0604	0.0286

Note:The coefficient reported in this table is the coefficient which captures the average treatment effect of the bilingual program on the dependent variable.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Triple Diff-in-Diff results for reading attitudes: estimated treatment effect by group

	Positives					
	Favourite hobbies	**Talk about books**	**Happy as present**	**Enjoy bookstore**	**Express opinions**	**Exchange**
Total	0.019	-0.104	0.384**	0.093	-0.035	-0.093
	(0.194)	(0.227)	(0.188)	(0.134)	(0.217)	(0.201)
Controls	Yes	Yes	Yes	Yes	Yes	
Obs. Students	1,510	1,510	1,510	1,510	1,510	1,510
Adjusted R-squared	0.0658	0.0636	0.0604	0.0755	0.0451	0.0853
	Negatives					
	Only if I have to	**Hard to finish**	**Waste of time**	**Need information**	**Cannot l sit still**	
Total	-0.117	-0.126	0.129	-0.109	0.128	
	(0.223)	(0.224)	(0.200)	(0.223)	(0.169)	
Controls	Yes	Yes	Yes	Yes	Yes	
Obs. Students	1,510	1,510	1,510	1,510	1,510	
Adjusted R-squared	0.0659	0.0135	0.0470	0.0679	0.0341	

Note:The coefficient reported in this table is the coefficient which captures the average treatment effect of the bilingual program on the dependent variable.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11 displays the estimates results of the bilingual program impact on high school perceptions. In the questionnaire a question asks about different high school climate sentences in which students can select the answers in the same spirit than reading habits [14]. I do analyze these questions by creating dummy variables as before. I do find relatively small negative and significant results on students' perception that high school classes improve students' confidence of 15.9 pp (19 percent of the baseline mean). This result is mostly driven by students whose parents do not have university education, which reports negative and significant results of 20 pp (22 percent of the baseline mean), whereas the rest are not affected. Moreover, students whose parents do not have college education report a negative and significant result of 18.4 pp (18.4 percent of the baseline mean) in the opinion that high schools are useful for finding a job. It seems that bilingual program affects negatively students' perception of high-school usefulness for finding a job and students' confidence. This negative effect is mainly for students whose parents do not report a university degree.

Table 11: Diff-in-Diff results for high school perceptions: estimated treatment effect by group

	Little preparation for Adult Life	Waste of Time	Improve Confidence	Useful for Jobs
Total	-0.099	0.007	-0.159*	-0.086
	(0.097)	(0.057)	(0.089)	(0.064)
Baseline mean	.28	.08	.84	.92
Controls	Yes	Yes	Yes	Yes
Obs. Students	1,510	1,510	1,510	1,510
Adjusted R-squared	0.0200	0.00726	0.0164	0.0116
University	-0.045	0.012	-0.141	-0.045
	(0.118)	(0.079)	(0.127)	(0.103)
Baseline mean	.21	.07	.79	.86
Controls	Yes	Yes	Yes	Yes
Obs. Students	760	760	760	760
Adjusted R-squared	0.0287	0.00411	0.0163	0.0231
Non-University	-0.146	0.027	-0.200*	-0.184***
	(0.152)	(0.082)	(0.119)	(0.055)
Baseline mean	.36	.09	.91	1
Controls	Yes	Yes	Yes	Yes
Obs. Students	750	750	750	750
Adjusted R-squared	0.00849	0.00878	0.00580	0.00531
Triple Diff-in-Diff	0.091	-0.049	0.040	0.146
	(0.192)	(0.111)	(0.170)	(0.115)
Controls	Yes	Yes	Yes	Yes
Obs. Students	1,510	1,510	1,510	1,510
Adjusted R-squared	0.0190	0.00792	0.0129	0.0121

Note:The coefficient reported in this table is the coefficient which captures the average treatment effect of the bilingual program on the dependent variable.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 12 exhibits the results of the bilingual program impact on teachers opinion.

[14]Question 09-ST33 in PISA 2009 and AP30 in 2013 data.

In the questionnaire a question asks about different sentences regarding high school teachers in which students can select the answers in the same spirit than reading habits [15]. I analyze these questions by creating dummy variables as before. I do not find statistically significant effects on teachers opinion. The bilingual program does not seem to improve or to hurt students' opinions of teachers for neither for all students nor by parental education.

Table 12: Diff-in-Diff results for teachers opinions: estimated treatment effect by group

	Get along well	**Interested in well-being**	**Really listen**	**Extra help**	**Treat me fairly**
Total	-0.045	0.102	-0.086	0.050	0.031
	(0.084)	(0.115)	(0.105)	(0.117)	(0.116)
Baseline mean	.8	.48	.72	.52	.6
Controls	Yes	Yes	Yes	Yes	Yes
Obs. Students	1,481	1,481	1,481	1,481	1,481
Adjusted R-squared	0.00667	0.0283	0.0225	0.00625	0.0156
University	0.030	0.029	-0.144	-0.116	-0.041
	(0.109)	(0.140)	(0.141)	(0.145)	(0.151)
Baseline mean	.78	.57	.78	.64	.64
Controls	Yes	Yes	Yes	Yes	Yes
Obs. Students	749	749	749	749	749
Adjusted R-squared	0.0162	0.0341	0.0272	0.00346	0.0239
Non-University	-0.117	0.221	0.001	0.234	0.109
	(0.113)	(0.174)	(0.160)	(0.171)	(0.173)
Baseline mean	.82	.36	.64	.36	.54
Controls	Yes	Yes	Yes	Yes	Yes
Obs. Students	732	732	732	732	732
Adjusted R-squared	-0.00322	0.0139	0.0124	0.00454	0.0170
Triple Diff-in-Diff	0.113	-0.151	-0.165	-0.326	-0.090
	(0.162)	(0.230)	(0.209)	(0.229)	(0.233)
Controls	Yes	Yes	Yes	Yes	Yes
Obs. Students	1,481	1,481	1,481	1,481	1,481
Adjusted R-squared	0.00651	0.0271	0.0239	0.00617	0.0175

Note:The coefficient reported in this table is the coefficient which captures the average treatment effect of the bilingual program on the dependent variable.

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[15]Question P09-ST34 in PISA 2009 and AP31 in 2013 data.

7 Concluding remarks

Using a unique survey of a 9th graders students sample from Madrid Region, I provide evidence about the effects of a public English-Spanish bilingual program on students' performance and other students' outcomes. Working with a Differences-in-Differences model with students' individual controls, I find no statistically significant effects of the program on Reading, Mathematics and Sciences performance at the high-school nor at the student level. The same is found for high school dispersion of performance. In contrast, I do find positive and significant effects on students' reading enjoying time and several beneficial reading habits (such as students' opinions of reading as a favorite hobby, happy when receiving a book as a present, and enjoying exchanging books), with higher impacts for students' whose parents have college education. Moreover, I find negative and significant results on students' perception that high school classes improve students' confidence, and an additional effect of students' perception that high schools are not useful for finding a job for children of parents with no college education. Finally, no statistically significant effects have been found on students' opinion of teachers .

The paper suggests the possibility of an evolution of the bilingual program effect on cognitive skills over time. The negative results on students' performance in primary education for children whose parents have less than upper secondary education found by Anghel et al. (2015) are consistent with the work developed by Bialystock (2007) who finds negative effects on literacy skills measured right after exposure to bilingual education. It contrasts with Admiraal et al (2006) and the present paper, finding no effects on secondary education. It seems that students experienced a negative effect right after the exposure to the bilingual program (primary education) and this ameliorated until no significant differences were found in secondary education. However, it may also be due to a selection effect on students, given the fact that having performed well in the bilingual program is a requirement to enroll in bilingual sections of high schools. Therefore, estimating the effect of the bilingual program for upper secondary education to test the long run effects of the program, as well as increasing the sample size in order to improve the statistical power of the regressions are potential avenues for further

research.

References

W. Admiraal, G. Westhoff, and K. de Bot. Evaluation of bilingual secondary education in the netherlands: Students' language proficiency in english. *Educational Research and Evaluation*, 12(1):75–93, 2006.

B. Anghel, A. Cabrales, and J. M. Carro. Evaluating a bilingual education program in spain: the impact beyond foreign language learning. 2015.

M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? Technical report, National Bureau of Economic Research, 2002.

E. Bialystok. Acquisition of literacy in bilingual children: A framework for research. *Language learning*, 57(s1):45–77, 2007.

L. Cappellari, A. Di Paolo, et al. Bilingual schooling and earnings. 2015.

A. Chin, N. M. Daysal, and S. A. Imberman. Impact of bilingual education programs on limited english proficient students and their peers: Regression discontinuity evidence from texas. *Journal of Public Economics*, 107:63–78, 2013.

A. Costa and N. Sebastián-Gallés. How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience*, 15(5):336–345, 2014.

J. Fidrmuc and J. Fidrmuc. Foreign languages and trade. 2009.

V. A. Ginsburgh and J. Prieto-Rodriguez. Returns to foreign languages of native workers in the european union. *Industrial & Labor Relations Review*, 64(3):599–618, 2011.

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

C. Jepsen. Bilingual education and english proficiency. *Education*, 5(2):200–227, 2010.

F. Kreuter and R. Valliant. A survey on survey statistics: What is done and can be done in stata. *Stata Journal*, 7(1):1, 2007.

K. D. Munshi and M. R. Rosenzweig. Traditional institutions meet the modern world: Caste, gender and schooling choice in a globalizing economy. 2003.

P. A. Patrician. Multiple imputation for missing data. *Research in Nursing & Health*, 25 (1):76–84, 2002.

C. H. Quecedo. The impact of bilingual education on average school performance: an evaluation of madridâĂŹs bilingual schools programme.

A. Saiz and E. Zoido. Listening to what the world says: Bilingualism and earnings in the united states. *Review of Economics and Statistics*, 87(3):523–538, 2005.

A. C. Willig. A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of educational research*, 55(3):269–317, 1985.