UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 12-14
Economic Series
May, 2012

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Evaluating a bilingual education program in Spain: the impact beyond foreign language learning*

Brindusa Anghel
FEDEA.
banghel@fedea.es

Antonio Cabrales
Universidad Carlos III de Madrid
acabrale@eco.uc3m.es

Jesús M. Carro
Universidad Carlos III de Madrid
jcarro@eco.uc3m.es

Abstract

We evaluate a program that introduced bilingual education in English and Spanish in primary education in some public schools of the Madrid region in 2004. Under this program students not only study English as a foreign language but also some subjects (at least Science, History and Geography) are taught in English. Spanish and Mathematics are taught only in Spanish. The first class receiving full treatment finished Primary education in June 2010 and they took the standardized test for all 6th grade students in Madrid on the skills considered "indispensable" at that age. This test is our measure of the outcome of primary education to evaluate the program. We have to face a double self-selection problem. One is caused by schools who decide to apply for the program, and a second one caused by students when choosing school. We take several routes to control for these selection problems. The main route to control for self-selected schools is to take advantage of the test being conducted in the same schools before and after the program was implemented in 6th grade. To control for students self-selection we combine the use of several observable characteristics (like parents' education and occupation) with the fact that most students were already enrolled at the different schools before the program was announced. Our results indicate that there is a clear negative effect on learning the subject taught in English for children whose parents have less than upper secondary education, and no clear effect for anyone on mathematical and reading skills, which were taught in Spanish.

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

* We gratefully acknowledge the help of José Carlos Gibaja and Ismael Sanz to obtain the data and the support from the Spanish Ministry of Science and Technology from grants ECO2009-07530 (Anghel), ECO2009-10531 (Cabrales) ECO2009-11165 (Carro), CONSOLIDER-INGENIO 2010 - CSD2006-0016 (all) and Consejería de Educación de la Comunidad de Madrid under the Excelecon project.

1 Introduction

Knowledge of a second language is widely believed to be essential for workers to succeed in an increasingly interconnected business world, and researchers tend to agree. Ginsburgh and Prieto-Rodríguez (2011), for example, found large estimates of the effects of foreign language knowledge on wages in Mincerian regressions: the increases in wages ranged between 11 percent in Austria and 39 percent in Spain for knowledge of the English language and even higher effects for knowledge of other languages.[1,2] The returns to learning English do not only flow to individuals, the country as a whole may also benefit: Fidrmuc and Fidrmurc (2009) show, for example, that widespread knowledge of languages is an important determinant for foreign trade, with English playing an especially important role.

The private initiative has taken notice of these benefits of second language acquisition. Many schools, in Spanish speaking countries especially those that cater to the elites, offer bilingual education for their pupils; Banfi and Day (2004) document this for Argentina, and Ordoñez (2004) for Colombia. The high returns for foreign language capabilities, and probably also the association with elite schools, have prompted several Spanish administrations to offer bilingual education in schools across the country. The ministry of education sponsors an agreement with the British Council that selects 80 schools all over Spain where instruction in English occupies a large percentage of the curriculum. Much more ambitious in scale is a program in the autonomous region of Madrid which at present enrolls 340 public schools (276 primary schools and 64 high schools)[3] where around 40 percent of the instruction, including all the science curriculum, is taught in English.[4] These programs have been so successful with voters that both major parties have included in their 2011 general election platforms the promise of extending the program to the whole nation.[5]

It is thus clear, both to researchers and the general public, that learning a foreign language is important for economic reasons. But it also has some costs. The more obvious are the financial ones: the teachers may need to be hired, trained, or retrained, and given the market value of English knowledge they will be more costly than other teachers; some

[1] An earlier analysis of the same data, by Williams (2011) found a smaller impact: between 5 percent in Austria or Finland, to insignificant in Spain or France. But the reanalysis of Ginsburgh and Prieto-Rodríguez (2011) used more powerful techniques to control for endogeneity.

[2] The effects on U.S. workers are rather smaller, as one would expect from the *lingua franca* status of English. See for example Fry and Lowell (2003) who find no effect on wages, or Saiz and Zoido (2005) who find an effect of about 5%.

[3] The 276 primary schools represent 35% of the total number of public schools and the 64 high schools represent 20% of the total number of high schools in the region of Madrid.

[4] Andalusia also has a bilingual program, but the percentage of instruction in English is smaller, around 20 percent of the instruction time.

[5] See e.g. in the program of the socialist party PSOE the statement "we will support the design of linguistic projects to support the learning of English. We will also support the schools offering bilingual education both in vocational training and at the university,"(available at: $http://www.rubalcaba.es/wp-content/uploads/2011.pdf/10/progpsoe2011.pdf$) or the one of conservative party PP, which states "We will promote Spanish-English bilingualism in the whole educational system from kindergarten to university", (available at: $http://www.pp.es/actualidad-noticia/programa-electoral-pp_5741.html$).

extra conversation assistants may need to be hired; if successful, demand will grow and the program may need to be expanded. But in addition to these costs time is finite, and there is hardly ever a free lunch in educational issues; so there may be other negative effects from the policy. The aim of this paper is precisely to test whether bilingual educational programs have a cost in terms of slower learning rates in other subjects.

To test this idea we look at data from the bilingual education program in the region of Madrid. Although we will describe it in more detail later, the program (for primary schools) basically consists on using English to teach the subject called "Knowledge of the Environment", that includes all teaching of Science, History and Geography. English is also used as the educational medium for Art and sometimes Physical Education, and of course the English language classes. Overall, teaching in English comprises between 10 and 12 of the 25 weekly hours of instruction.

To find out the effects of the program we use a standardized exam that has been administered each year in all primary schools from the Spanish region of Madrid to 6th grade students (12-13 years of age), starting with the school year 2004/05. The exam tests for what are called "Indispensable Knowledge and Skills" in three areas: Spanish language, Mathematics and General Knowledge; the latter basically corresponds to the material taught in "Knowledge of the Environment". The exam results are anonymous, but each student answers a questionnaire that includes a host of socioeconomic background variables, which we can use as covariates. We use data from the first group of schools that became bilingual in the region of Madrid in 2004/05, and we checked the results of the student cohorts which took the exam in 2009/10 and in 2010/11. We then repeat the analysis with the second group of schools that became bilingual for their first bilingual cohort, whose students took the exam in 2010/11. In order to control for endogeneity problems, we use a *Difference in Difference* approach, comparing the exam results of children in the treated schools before and after they became bilingual with the group of non-bilingual schools before and after the treatment.

We find that the effect of the program is not significantly different from zero for either Mathematics or Spanish language, although it goes from positive to negative. For General Knowledge, the bilingual program has a negative and significant effect on the exam results, for children of parents without a college education. The size of this effect is substantial, on the order of 0.2 standard deviations.[6] Since General Knowledge is the only subject taught in English from the three present in the exam, it would appear that the extra effort made to use English as the medium of instruction comes at the expense of a worsening in the learning of that subject. A possible *caveat* to that conclusion is that the exam is taken in Spanish and the subject is learnt in English. But, taken at face value, this would also suggest that the level of linguistic competence in English is not enough to leap through that barrier. All in all, the conclusion must be that there is indeed no free lunch: either

[6] This is close in magnitude to the effects found by Angrist and Lavy (1999) in Israel for a class reduction of 8 students, and by Krueger (1999) for the Tennessee STAR experiment, which reduced class size in 7 students.

the learning of subjects taught in English is impaired, or the learning of English itself is not very good.

In the group of schools that started to participate in 2004 the results for the second cohort of students exposed to the program are very similar, even quantitatively, to those of the first cohort. However, for the group of schools that started to participate in 2005, the effects are also negative and significant only for General Knowledge, but they are smaller in size and only for children of parents with less than upper secondary education. We conjecture that this is due to a better selection of those schools in terms of the English knowledge of the teachers, since for that group of schools the conditions to be a part of the program were made stricter in that dimension.

There is a large body of research aimed at understanding the effects of bilingual education programs for immigrants in the U.S. This literature finds mostly positive results of those programs. Willig (1985) concludes that the better the experimental design of the study, the more positive were the effects of bilingual education, and Greene (1998) in another meta-study of the literature asserts that: "an unbiased reading of the scholarly research suggests that bilingual education helps children who are learning English." Jepsen (2009), on the other hand finds that "students in bilingual education have substantially lower English proficiency than other English Learners in first and second grades. In contrast, there is little difference between bilingual education and other programs for students in grades three through five." But those are typically programs for immigrants into a foreign country so the external validity to our population of those results is rather unclear.

There is much less evidence regarding the effects of bilingual education in English for countries whose official language is not English. An exception is Admiraal, Westhoff and de Bot (2006), who study the effect of the use of English as the language of instruction for secondary education in The Netherlands. They state that: "No effects have been found for receptive word knowledge and no negative effects have been found with respect to the results of their school leaving exams at the end of secondary education for Dutch and subject matters taught through English." It is hard to know what to make of the differences between our two studies, since the educational systems are very different, as are the societies where the programs are administered. But an intriguing question arises: could the costs of bilingual education be lowered if the program was started in high school? This is an important question for further research.

The paper is organized as follows. Section 2 describes in some detail the institutional setup and the program. Section 3 discusses the data and the econometric model. Section 4 contains the main results of the paper and it has some additional estimations and robustness checks. Section 5 concludes.

2 Institutional Background and Description of the Program

The order from the regional ministry of Madrid that initiated the bilingual school program argues that it is needed because: "The full integration of Spain in the European context implies that students need to acquire more and better communication skills in different European languages. Being able to develop their daily and professional activities using English as a second language opens new perspectives and new relationship possibilities to students of bilingual schools in the Autonomous Region of Madrid." The integrated European labor and trading market is thus the reason used by the administration for fostering the program.

This is a good reasoning, in the current recession with a general unemployment rate above 20 percent and a youth unemployment rate of almost 50 percent, only 36,967 Spaniards emigrated in 2010. This contrasts markedly with the 4 million unemployed, or with the 40,000 yearly emigrants that Bergin et al. (2009) estimate for Ireland, a country 10 times smaller than Spain and with half its unemployment rate. Of course, there are many reasons for this, Bentolila and Ichino (2008) argue that the welfare state and the family make it possible to accommodate big unemployment shocks, but the welfare state and the family are similar in Spain and Ireland, so it is indeed quite likely that the lack of proficiency of adult Spanish cohorts in English is one problem hindering the emigration that the unemployment figures would suggest should be a safety valve for the situation.

The Spanish educational system is composed of 6 years of primary school, 4 years of compulsory secondary education (E.S.O.) and 2 years of non-compulsory education, which is divided into vocational training (ciclos formativos) and preparation for college (bachillerato). There are also three years of free publicly funded pre-school, from ages 3 to 5. The pre-school children share the premises with those in primary school. Also, the pre-schoolers in one location have precedence over other children applying to the same primary school. As a consequence of this precedence rule most students at the primary level come from the preschool in the same location. In fact, if all the vacancies for three years old are filled and none of them leaves the school at the primary level, there will not be any vacancies at that level in that cohort. As a result, the school choice is almost universally made when the student is three years old. After that time, school changes are not frequent, because it becomes extremely difficult to enter schools with high demand.

The facts mentioned about school choice and selection in the previous paragraph are important for our study. The bilingual program is applied at the primary school level, not at pre-school. Since at the time the bilingual program was designed and announced there were students already in the pre-school level at the selected schools, their parents' school choices were made three years prior to that moment, when the program did not exist and was not even planned. For this reason the differences between the first cohort of treated students and the previous cohorts cannot be related to the introduction of the program.

The program started with children in the first grade of the selected primary schools in the school year 2004/05 and left others in the same school, and all in the remaining schools, untreated. The program progressed with their school training for those treated students. Successive cohorts from the treated schools have also been treated, and additional primary schools joined the program in successive years, always starting the treatment with first graders. Our data covers only the schools from the first cohort. Once the students from the 2004/05 cohort reached secondary education (in 2010/11), a second phase kicked in and some high schools joined the program. Since that phase of the program is still in progress, we will not be able to analyze it.

The program was initiated in 2004 with a call for applications by schools, of which 25 were selected in the first year[7], with initial plans for extension up to 110, which were later expanded to the present 276 due to the high demand (out of a total of about 740 public schools). A school wishing to be selected for the program had to submit an application. The three criteria used to evaluate those applications are:

1. *Degree of acceptance of the educational community* expressed through the support received by the application by the school teachers and the School Board (a decision making body composed of the principal and elected teachers and parents).

2. *Feasibility of the application.* This will take into account the previous experience of the school (some schools had started small pilot programs on their own), teaching staff, particularly the teachers with an English specialization, the school resources and the number of classes and students.

3. *Balanced distribution of selected schools between the different geographical areas*, taking into account the school population between three and sixteen.

The selected schools were not the 25 that best meet the first two criteria because of the criterion for geographical equity. However, the selected schools had all close to top grades in those criteria.

For the schools that were selected into the program in the following years, from 2005 onwards, the criteria used in the evaluation changed in one significant way. The former rule *3.* was replaced by

3'. *English level of the teachers in the school.* This level is verified either with some official certificate (such as those awarded by the University of Cambridge) that accredits a sufficient level of command of the English language or by an evaluation done directly by the education department of the regional government.

The balanced distribution is still mentioned as a desirable property of the allocation but it is not given explicit points.

The order calls bilingual a school where the language of instruction is English during at least one third of the school time, and where English language classes take 5 weekly periods (of 45 to 60 minutes). It explicitly excludes the Spanish language and Mathematics classes from being taught in English.

In Table 1 we describe the weekly curriculum from first to sixth grade so that it becomes clear the margin of autonomy in the number of teaching hours.

[7] In fact, there were 26 schools that became bilingual in 2004/05, out of which we have enough information on 25 schools.

Table 1: Weekly schedule by area in primary school

	Number of weekly hours					
	First cycle		Second cycle		Third cycle	
Areas	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade	6th Grade
Know. Environ.	4	4	4	4	4	4
Art	3	3	2.5	2.5	2.5	2.5
Physical Educ.	3	3	3	3	2.5	2.5
Spanish Language	5	5	5	5	5	5
Foreign Language	2	2	2.5	2.5	3	3
Mathematics	4	4	4	4	4	4
Culture, religion	1.5	1.5	1.5	1.5	1.5	1.5
Recess	2.5	2.5	2.5	2.5	2.5	2.5
Total	25	25	25	25	25	25

With Knowledge of the Environment (a subject encompassing science, geography and history) plus 5 periods of English, the minimum is accomplished. Different schools choose whether to increase the English instruction by also teaching in that language Art, Physical Education and Religion (or its alternative for those not wanting Religion, which is mostly a class in social norms and culture). Whether English instruction is expanded from the minimum depends on the availability of teachers, but most schools end up having above 40 percent of the instruction in English.

The program is certainly not costless. The teachers involved in it receive a complement over their basic wage based on 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 regular attendance at coordination meetings outside school hours." The extra work is estimated by the order to be "on average of three hours per week for teachers, and four hours for coordinators." The order does not say how the administration arrived at this estimate. To compensate for the extra dedication the coordinators of the program in each school receive 1,980 euros a year; a teacher who teaches more than 15 hours in English, for subjects different than English language, 1,500 euros; between 8 and 15 hours, 1,125 euros; and less than 8 hours, 750 euros. The program provides "conversation assistants" to schools, typically college students from English speaking countries. Finally, the program provides training courses in English for teachers, both in Spain and abroad. In the latter case, the program covers transportation, living expenses and fees for English schools, mostly in the UK and Ireland.

In order to teach in English, the teachers have to be either specialists in English or pass an exam. The exam is divided in two parts. The first part is a written exam, where they are tested on reading, writing and listening comprehension, plus vocabulary and grammar. The second part is oral and involves a 20 minutes conversation with the examiner.

3 Description of Data and Econometric Model

3.1 Description of Data

Our data comes from a standardized exam that has been administered each year in all primary schools from the Spanish region of Madrid to 6th grade students (12-13 years of age), starting with the school year 2004/05.[8] The exam is called CDI *(prueba de Conocimientos y Destrezas Indispensables)*, which means "Indispensable Knowledge and Skills Exam". It is compulsory for all schools (public, private or charter). Like the OECD's PISA exam, the CDI exam does not have any academic consequences for the student, it is only intended to give additional information to teachers, parents and students.

The exam consists of two parts of 45 minutes each: the first part includes tests of Dictation, Reading, Language and General Knowledge and the second part is composed of mathematics exercises. We use as a measure of student achievement the exam scores, standardized to the yearly mean, in General Knowledge (whose contents are close to the subject "Knowledge of the Environment" which is taught in English) and in Reading and Mathematics (which are taught in Spanish). The exams are conducted in Spanish for all students, whether or not they were in a bilingual school.

Before taking the exam, a short questionnaire (see Appendix) is filled out by each student. In the questionnaire the students are asked a few questions about themselves, their parents and the environment in which they are living. The answers to these questions provide rich information on individual characteristics of students: from the questionnaire we obtain the age of the student; the country of birth, which we divide into Spain, China, Latin America, Morocco, Romania and other, to have sufficiently many observations of each category; the level of education of the parents; the occupation of the parents; the composition of the household in which the students lives; and the age at which the student started to go to school/kindergarten. From the exam we have information at student level on gender, whether the student has any special educational needs and whether the student has any disability.

Regarding the education of the parents, students were asked to provide this information for both the mother and the father. In order to facilitate the interpretation we choose the highest level of education between the mother and the father. We distinguish the following categories: university education, higher secondary education, vocational training, lower secondary education and no compulsory education. The same applies to the occupation of the parents: since we have the occupation of both the mother and the father, we choose the highest level between them. Thus, we differentiate between the following categories: professional occupations (for example teacher, researcher, doctor, engineer, lawyer, psychologist, artist, etc.); business and administrative occupations (for example CEO, civil servant, etc.); and blue collar occupations (for example shop assistant, fireman,

[8] Since the school year 2009/10 the exam is also administered to all students in the third grade of compulsory secondary education (14-15 years old).

construction worker, cleaning staff, etc.).[9]

The variable on the composition of the household of the student comes from the answers to the question: "With whom do you usually live?". We differentiate the following seven categories: lives only with the mother, lives with the mother and one sibling, lives with the mother and more than one sibling, lives with the mother and the father, lives with the mother and the father and one sibling, lives with the mother and the father and more than one sibling and other situations.

For our empirical analysis we use data of the first cohort of bilingual schools in the region of Madrid which started first grade of primary school in 2004/05, and took the CDI exam in 2009/10 (25 schools).

In order to control for the endogeneity problems caused by self-selection of students and schools which we will explain below, we use a Difference in Difference approach. We compare the performance of children in the treated schools before and after they became bilingual with the group of non-bilingual schools before and after the treatment. Thus, we employ the data for 2008/09 and 2009/10 cohorts. The four groups that we analyze are the following: the group of bilingual schools in 2008/09 (the treatment group before the change), the group of non-bilingual schools in 2008/09 (the control group before the change), the group of bilingual schools in 2009/10 (the treatment group after the change) and the group of non-bilingual schools in 2009/10 (the control group after the change).

Table 2 provides descriptive statistics of these four groups. If we compare the schools where the bilingual program was introduced, before and after the treatment, we see an increase in the proportion of students with characteristics that are positively correlated with academic performance. More concretely, the proportion of children whose parents have university education increases from 33% to 39%, the proportion of children whose parents have lower secondary education decreases from 26% to 22% and the proportion of children whose parents did not finish compulsory studies also decreases from 8% to 5%. There are also important changes with regards to the occupations of the parents of children from these two cohorts: the proportion of children whose parents have professional occupations increases from 24% to 29% and the proportion of children whose parents have blue collar occupations decreases from 58% to 51%.

Furthermore, there is an increase in the proportion of Spanish students from the 2008/09 group to the 2009/10 group from 81% to 87%, which translates in a decrease in the proportion of immigrant students (the most important change is in the reduction of the proportion of Latin American students from 10% to 6%, whose performance is generally worse than that of Spanish students or even of other immigrants, after conditioning on observables (Anghel and Cabrales, 2010)). We also detect an increase in the percentage of children who started school before 3 years from 46% to 51%.

However, if we look at the control group we do not see any important changes in the composition of cohorts from one year to another: these proportions remain almost constant in both years (at most there is a difference of one decimal).

[9] Robustness checks using separately the education of each parent yield very similar results.

Table 2: Descriptive statistics benchmark

	Treat. bef.	**Cont. bef.**	**Treat. aft.**	**Cont. aft.**	Diff-in-Diff
Variable	Mean	Mean	Mean	Mean	
Subjects					
Dictation	5.29	5.59	7.90	7.89	0.31
Mathematics	8.94	9.54	10.55	10.88	0.26
Language	10.44	10.87	14.60	14.84	0.18
Reading	2.87	2.93	3.53	3.59	0.01
General knowledge	2.28	2.35	3.17	3.37	-0.13
Subjects - standard.					0.00
Dictation	-0.09	0.00	0.00	0.00	0.09
Mathematics	-0.11	0.00	-0.06	0.00	0.05
Language	-0.08	0.00	-0.05	0.00	0.03
Reading	-0.04	0.00	-0.04	0.00	0.00
General knowledge	-0.05	0.00	-0.15	0.00	-0.11
Individual charac.					
Female	0.50	0.49	0.51	0.49	0.01
Stud. with special ed.	0.11	0.07	0.06	0.06	-0.04
Student with disab.	0.04	0.03	0.03	0.03	-0.01
Student's age	12.15	12.14	12.12	12.14	-0.04
Student Spain	0.81	0.81	0.87	0.81	0.06
Student Romania	0.03	0.02	0.02	0.02	-0.01
Student Morocco	0.01	0.01	0.00	0.01	0.00
Student Lat.Am.	0.10	0.11	0.06	0.10	-0.03
Student China	0.00	0.01	0.00	0.01	0.00
Student other	0.05	0.04	0.04	0.05	-0.01
Parent education					
Univ.	0.33	0.48	0.39	0.47	0.07
Higher secondary	0.21	0.17	0.20	0.18	-0.02
Vocational training	0.12	0.12	0.14	0.12	0.01
Lower secondary	0.26	0.17	0.22	0.17	-0.04
Did not finish comp.	0.08	0.06	0.05	0.05	-0.02
Parent profession					
Business, civil serv.	0.17	0.22	0.19	0.22	0.02
Professional	0.24	0.33	0.29	0.33	0.05
Blue Collar	0.58	0.46	0.51	0.45	-0.06
Age start. sch.					
Start school before 3	0.46	0.51	0.51	0.54	0.02
Pre-school 3-5	0.49	0.44	0.47	0.43	0.00
Start school at 6	0.03	0.03	0.02	0.02	-0.01
Start sch. after 6	0.02	0.01	0.01	0.01	-0.01
Obs. Schools	25	1201	25	1217	
Obs. Students	1135	55793	1145	53150	

The numbers presented above suggest that there could be an endogenous change in the characteristics of the students enrolled in the bilingual schools, before and after the treatment. This change involves an improvement in student characteristics like the level of education and the occupation of parents or their nationality, which are known to be determinants of the academic performance of children.[10]

Moreover, the change in observable characteristics from one year to the next suggests that, apart from the treatment, there could be a change in unobservable characteristics. To find out whether this is the case, we analyze further data about the students in these bilingual schools.

A possible explanation for the changes between cohorts could come from students who entered in, or dropped out of, these schools after they became bilingual. These flows of students could generate some of the changes we observe. To check this theory, we obtained the list of children who attended the treated schools since they were five years old, the last year of pre-school education.

With that list, first, we analyze the group of schools where the number of children who entered after they became bilingual (that is, children who were not enrolled in that school when they were 5 years old) is less than 4 (that is about 16 percent in the average class of 25). We consider these schools as schools with a small number of incoming students, and the socioeconomic composition of the cohorts should not vary much from one year to the next one. There are eight treated schools that satisfy this condition. As before, we compare these schools before they became bilingual (the 2008/09 cohort) and after they became bilingual (the 2009/10 cohort) and we use as a control group the group of non-bilingual schools (we drop from the descriptive statistics the other 17 bilingual schools).

The descriptive analysis in Table 3 shows a very similar picture to the one in Table 2. We see that the change in the characteristics of students from the year in which they became bilingual to the next one goes in the same direction and is quantitatively similar as for the whole sample. We observe an important increase in the proportion of students whose parents have university degrees, from 27% in the 2008/09 cohort to 36% in the 2009/10 cohort, and a decrease in the proportion of students whose parents did not finish compulsory education (from 8% to 5%). We also identify a small increase in the proportion of students whose parents have professional occupations and a small drop in the proportion of students whose parents have blue collar occupations. Furthermore, there is an increase in the proportion of Spanish students from one cohort to the next one in the treated schools and there is a big drop in the proportion of Latin American students. Finally, the percentage of children who started to go to kindergarten before three years old increases by six percentage points (from 44% to 50%). Altogether, the selection problem that we detected with the full sample persists in the sample of eight schools with very few incoming students after they became bilingual.

Second, we restrict further the group of students we analyze, by studying only the characteristics of the group of children that were already enrolled in the 25 treated schools

[10] In the case of Madrid and for this same CDI exam this is shown in Anghel and Cabrales (2010).

Table 3: Descriptive statistics Schools with few movements

	8 sch. before	**8 sch. after**
Variable	Mean	Mean
Subjects		
Dictation	5.46	7.97
Mathematics	8.73	10.48
Language	10.65	14.68
Reading	2.92	3.61
General knowledge	2.28	3.11
Subjects - standard.		
Dictation	-0.04	0.03
Mathematics	-0.15	-0.07
Language	-0.04	-0.04
Reading	-0.01	0.02
General knowledge	-0.05	-0.20
Individual charac.		
Female	0.49	0.50
Stud. with special ed.	0.08	0.07
Student with disab.	0.05	0.04
Student's age	12.17	12.12
Student Spain	0.85	0.93
Student Romania	0.02	0.01
Student Morocco	0.01	0.00
Student Lat.Am.	0.10	0.05
Student China	0.00	0.00
Student other	0.03	0.01
Parent education		
Univ.	0.27	0.36
Higher secondary	0.20	0.22
Vocational training	0.15	0.12
Lower secondary	0.31	0.25
Did not finish comp.	0.08	0.05
Parent profession		
Business, civil serv.	0.17	0.20
Professional	0.23	0.26
Blue Collar	0.60	0.54
Age start. sch.		
Start school before 3	0.46	0.55
Pre-school 3-5	0.52	0.44
Start school at 6	0.02	0.01
Start sch. after 6	0.01	0.00
Obs. Schools	8	8
Obs. Students	416	434

since they were five years of age and started the bilingual education program in these schools. The introduction of the bilingual education program was not announced in advance of enrolling those children in the treated schools, so there should not be any changes in the characteristics of the treated children endogenous to the treatment. This analysis produces almost identical conclusions as in the previous cases (Table 4): we detect an increase in the proportion of students with characteristics that are positively correlated with their academic performance and this fact reveals once again a selection problem.

Third, we analyze the group of new incoming children in the 25 schools that became bilingual in 2004/05, in order to see whether their demographic characteristics could be a partial source of endogeneity.

From Table 5 it is clear that these students have a socio-economic background which is very similar to the one of the remaining students of the bilingual schools. There is only one exception; it looks like the proportion of immigrant students among the new incoming students is significantly higher: about 29% of the new incoming students are immigrants (out of which 12% are Latin Americans) while only 13% of all students in the bilingual schools are immigrants (out of which 6% are Latin American).

Finally, we examine the sample of schools that applied unsuccessfully to the call for the bilingual education program, and whose score was very close to the cut-off for being part of the program. There are 38 schools that satisfy these conditions. If these schools are similar to the schools that became part of the program, they would represent a better control group than the whole group of schools. In addition, if we see for those schools a similar change in demographics from one year to the next one as the change that we see for our treated group, this could indicate that the explanation for this change does not necessarily lie in the introduction of the bilingual education program.

The descriptive statistics of these schools in Table 6 reveal that both hypotheses are partially valid. First, these schools are more similar in demographics to the treated bilingual schools than to the schools from the complete control group (comparison with column 3 from Table 2). However, there are differences: the most important difference is that the proportion of Latin American students in this new group of schools is bigger than in the bilingual schools. Secondly, the characteristics of children change from the 2008/09 cohort to the 2009/10 cohort in the same direction as they change for the bilingual schools for those cohorts, even though these changes are a bit smaller than in the bilingual schools.

There is one striking phenomenon regarding this group of schools. The average scores of their students are significantly lower than the scores of the students of the bilingual schools in the year before the treatment (2008/09). However, in the 2009/10 CDI exam, the scores of the students in these schools improve considerably, reaching almost the same levels as the scores of the students in the bilingual schools from 2009/10.

Nevertheless, given the similarities between this group of schools and the treated schools, in the next section, as a robustness check, we will use this group of schools as a control group.

These descriptive analyses show that there has been an important change in the com-

Table 4: Descriptive statistics children who did not move

	Treat. Bef.	**Cont. Bef.**	**Treat. Aft.**	**Cont. Aft.**	**Diff-in-Diff**
Variable	Mean	Mean	Mean	Mean	
Subjects					
Dictation	5.29	5.59	8.04	7.89	-0.45
Mathematics	8.94	9.54	10.54	10.88	-0.25
Language	10.44	10.87	14.76	14.84	-0.35
Reading	2.87	2.93	3.57	3.59	-0.05
General knowledge	2.28	2.35	3.16	3.37	0.14
Subjects - standard.					
Dictation	-0.09	0.00	0.05	0.00	-0.14
Mathematics	-0.11	0.00	-0.06	0.00	-0.05
Language	-0.08	0.00	-0.02	0.00	-0.06
Reading	-0.04	0.00	-0.01	0.00	-0.03
General knowledge	-0.05	0.00	-0.16	0.00	0.12
Individual charac.					0.00
Female	0.50	0.49	0.51	0.51	0.02
Stud. with special ed.	0.11	0.07	0.05	0.07	0.06
Student with disab.	0.04	0.03	0.04	0.03	0.01
Student's age	12.15	12.14	12.09	12.15	0.07
Student Spain	0.81	0.81	0.93	0.81	-0.11
Student Romania	0.03	0.02	0.01	0.02	0.02
Student Morocco	0.01	0.01	0.01	0.01	0.00
Student Lat.Am.	0.10	0.11	0.04	0.10	0.06
Student China	0.00	0.01	0.00	0.01	0.00
Student other	0.05	0.04	0.02	0.05	0.03
Parent education					
Univ.	0.33	0.48	0.38	0.47	-0.05
Higher secondary	0.21	0.17	0.20	0.18	0.02
Vocational training	0.12	0.12	0.14	0.12	-0.01
Lower secondary	0.26	0.17	0.24	0.17	0.02
Did not finish comp.	0.08	0.06	0.05	0.05	0.02
Parent profession					
Business, civil serv.	0.17	0.22	0.20	0.22	-0.02
Professional	0.24	0.33	0.27	0.33	-0.02
Blue Collar	0.58	0.46	0.53	0.45	0.05
Age start. sch.					
Start school before 3	0.46	0.51	0.52	0.54	-0.03
Pre-school 3-5	0.49	0.44	0.47	0.43	0.00
Start school at 6	0.03	0.03	0.01	0.02	0.02
Start sch. after 6	0.02	0.01	0.00	0.01	0.01
Obs. Schools.	25	1201	25	1217	
Obs. Students	1135	55973	849	53150	

Table 5: Descriptive statistics children who moved

Variable	Mean	Std. Dev.
Subjects		
Dictation	7.55	2.99
Mathematics	10.62	5.86
Language	14.23	4.83
Reading	3.42	1.50
General knowledge	3.26	1.24
Subjects - standard.		
Dictation	-0.13	1.10
Mathematics	-0.05	1.07
Language	-0.14	1.09
Reading	-0.11	1.05
General knowledge	-0.08	0.98
Individual charac.		
Female	0.49	0.50
Stud. with special ed.	0.12	0.33
Student with disab.	0.03	0.16
Student's age	12.21	0.45
Student Spain	0.71	0.46
Student Romania	0.05	0.21
Student Morocco	0.02	0.14
Student Lat.Am.	0.12	0.33
Student China	0.01	0.09
Student other	0.10	0.30
Parent education		
Univ.	0.44	0.50
Higher secondary	0.19	0.39
Vocational training	0.13	0.34
Lower secondary	0.18	0.38
Did not finish comp.	0.06	0.25
Parent profession		
Business, civil serv.	0.20	0.40
Professional	0.35	0.48
Blue Collar	0.45	0.50
Age start. sch.		
Start school before 3	0.47	0.50
Pre-school 3-5	0.46	0.50
Start school at 6	0.05	0.22
Start sch. after 6	0.02	0.14
Obs. Schools	26	
Obs. Students	341	

Table 6: Descriptive statistics - Schools that applied to become a bilingual school and scored high in the selection criteria

Variable	Mean in CDI exam 2008/09	Mean in CDI exam 2009/10
Subjects		
Dictation	4.79	7.62
Mathematics	8.32	10.31
Language	9.32	14.47
Reading	2.46	3.51
General knowledge	2.06	3.34
Subjects - standard.		
Dictation	-0.23	-0.10
Mathematics	-0.22	-0.10
Language	-0.29	-0.08
Reading	-0.32	-0.06
General knowledge	-0.20	-0.02
Individual charac.		
Female	0.47	0.47
Stud. with special ed.	0.09	0.09
Student with disab.	0.04	0.05
Student's age	12.20	12.18
Student Spain	0.71	0.72
Student Romania	0.04	0.04
Student Morocco	0.01	0.02
Student Lat.Am.	0.17	0.16
Student China	0.00	0.01
Student other	0.06	0.06
Parent education		
Univ.	0.38	0.39
Higher secondary	0.20	0.21
Vocational training	0.11	0.11
Lower secondary	0.21	0.21
Did not finish comp.	0.10	0.07
Parent profession		
Business, civil serv.	0.19	0.17
Professional	0.22	0.27
Blue Collar	0.59	0.56
Age start. sch.		
Start school before 3	0.46	0.52
Pre-school 3-5	0.49	0.44
Start school at 6	0.03	0.02
Start sch. after 6	0.02	0.02
Obs. Schools	38	38
Obs. Students	1341	1292

position of the bilingual schools once they became bilingual, while in the control group of non-bilingual schools we do not see such differences. The self-selection problem that we possibly face in the case of the bilingual schools could contaminate our estimates, therefore we need to use econometric techniques that mitigate that problem.

3.2 Econometric model of education production

3.2.1 Model and endogeneity problems

Here we use as the outcome for primary education the *standardized* scores of students in the CDI exam described in section 3.1. For a given year, the score in that test for student i in school j, y_{ij}, is determined by:

$$y_{ij} = \delta bil_j + \beta x_i + v_j + u_i + \xi_{ij} \tag{1}$$

where x_i are the observable characteristics of students and their families described in section 3.1, bil_j indicates whether school j participated in the bilingual program, u_i are unobservable characteristics of the students, such as effort or ability, v_j are characteristics of the school, like quality of the Principal and teachers, and ξ_{ij} is a random shock. Our parameter of interest is the average effect of the bilingual program on y_{ij}, which in equation (1) is captured by δ. The difficulty that we face when we run the regression of y_{ij} on bil_j and x_i is that we could suffer from an endogeneity bias because of two self-selection problems:

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

2. *Schools are not randomly selected to implement the bilingual program.* The program was implemented only in (some of the) schools that applied for it. An application could be a positive signal of quality of the principal and teachers, because of the significant amount of extra work required by the program. It could also be a sign that the school had low demand (perhaps due to low quality) with teachers about to be displaced.[11]

3.2.2 Estimation strategy

To control for the endogeneity problem caused by the self-selection of schools and students explained, we use Difference in Differences estimation (diff-in-diff). This solves the self-selection of schools into the program because we observe the same school the first year

[11] In Spain a large majority of teachers are civil servants and cannot be fired. But they can be moved between schools within a region. Even in a small region like Madrid, this can entail substantial inconvenience and they would be willing to do significant efforts to avoid school closures.

the bilingual program is implemented in sixth grade and the year before. Given the institutional framework, the only significant changes in resources and staff from one year to the next are those associated with the bilingual program.

With respect to the self-selection of students, the diff-in-diff strategy also helps to solve this problem. As we mentioned in section 2 since the admission rules to primary school gives precedence to pre-schoolers in that same school, and given the timing of announcement of the program, the differences between the first cohort of treated students and the previous cohorts cannot be related to the introduction of the program. Given this observation, if the movements of students in bilingual schools after the program was introduced were the same as in the absence of the program (i.e. the same changes as in non treated schools) a diff-in-diff strategy would control for the students being differently distributed between treated and untreated schools. However, as one can see in Table 2 and we discussed in section 3.1, there is a change in the characteristics of the students in bilingual schools after the program was introduced. Fortunately the diff-in-diff easily allows us to incorporate observable characteristics of students in the estimation to control for this changes.

Given the diff-in-diff strategy, we are going to estimate the following regressions by OLS:

$$y_{ij} = \alpha_0 + a_1 bil_j + \alpha_2 y10 + \delta y10 \cdot bil_j + \varepsilon_{ij} \tag{2}$$

$$y_{ij} = \alpha_0 + a_1 bil_j + \alpha_2 y10 + \delta y10 \cdot bil_j + \beta x_i + \varepsilon_{ij} \tag{3}$$

where $y10$ is a dummy variable for the academic year 2009/10, the first year when we observe the children exposed to the bilingual education program in the CDI exam. Also, we will study further whether the change in the student population in bilingual schools is affecting our estimates by checking the robustness of our results to other comparisons and ways of estimating the effect of the program.

4 Results

4.1 Main estimates

In Table 7 we present estimates of models (2) and (3). The parameter associated with the variable Bilingual school 2004/05 in CDI exam 2009/10 ($y10 \cdot bil_j$) gives the effect of the program we want to estimate. Without covariates the effect of the program is not significant for the three subjects. However, as we mentioned when presenting the descriptive statistics of the data, the cohort of treated students has different characteristics than the previous cohort in those schools. Those characteristics affect positively the outcome; that is why the effect of the program is smaller once this change in observables is taken into account. This change in the estimated effect of the program when introducing covariates reflects the fact that there is selection in students after introducing the program. For mathematics and reading the effect is not significantly different from zero in either case,

Table 7: Diff-in-Diff with and without covariates. All students in the sample.

	Mathematics		Reading		General Knowledge	
Constant	0.002	4.517***	0.001	3.093***	0.001	3.391***
	(0.015)	(0.132)	(0.014)	(0.132)	(0.014)	(0.137)
Year 2010	-0.001	-0.073***	0.000	-0.084***	0.002	-0.072***
	(0.012)	(0.011)	(0.013)	(0.012)	(0.015)	(0.015)
Bilingual school 2004/05	-0.110	-0.006	-0.043	0.053	-0.046	0.069
	(0.074)	(0.058)	(0.096)	(0.091)	(0.093)	(0.094)
Bilingual school 2004/05	**0.053**	**-0.068**	**0.002**	**-0.110**	**-0.096**	**-0.229****
in CDI exam 2009/10	(0.075)	(0.069)	(0.096)	(0.099)	(0.102)	(0.112)
Female		-0.157***		-0.035***		-0.176***
		(0.007)		(0.006)		(0.007)
Student with special		-0.744***		-0.702***		-0.620***
educational needs		(0.017)		(0.019)		(0.020)
Student with disability		-1.080***		-1.127***		-0.892***
		(0.020)		(0.026)		(0.025)
Student's age		-0.384***		-0.262***		-0.280***
		(0.011)		(0.010)		(0.011)
Student Romania		0.036		0.017		0.061*
		(0.027)		(0.025)		(0.031)
Student Morocco		-0.053*		-0.256***		-0.147***
		(0.032)		(0.038)		(0.043)
Student Latin America		-0.249***		-0.073***		-0.193***
		(0.015)		(0.014)		(0.016)
Student China		0.600***		-0.282***		-0.319***
		(0.051)		(0.054)		(0.052)
Student other		-0.129***		-0.031**		-0.100***
		(0.017)		(0.016)		(0.016)
Parent education - Univ.		0.340***		0.273***		0.249***
		(0.016)		(0.018)		(0.018)
Parent education -		0.182***		0.173***		0.169***
Higher secondary		(0.015)		(0.018)		(0.017)
Parent education -		0.181***		0.204***		0.184***
Vocational training		(0.016)		(0.019)		(0.018)
Parent education -		0.100***		0.105***		0.102***
Lower secondary		(0.015)		(0.019)		(0.017)
Parent occupation -		0.167***		0.139***		0.102***
Business, minister, city hall		(0.010)		(0.010)		(0.011)
Parent occupation-		0.251***		0.205***		0.151***
Professional		(0.009)		(0.009)		(0.010)
Lives only with the mother		-0.099***		-0.080***		-0.079***
		(0.023)		(0.024)		(0.027)
Lives with the mother		0.071***		0.034		0.030
and one sibling		(0.025)		(0.025)		(0.029)
Lives with both parents		0.066***		0.003		0.065**
		(0.022)		(0.023)		(0.026)

continue in next page

Table 7: Diff-in-Diff with and without covariates. All students in the sample (cont.)

	Mathematics		**Reading**		**General Knowledge**	
Lives with both parents		0.174***		0.068***		0.100***
and one sibling		(0.022)		(0.022)		(0.025)
Lives with both parents		0.151***		0.055**		0.063**
and more than one sibling		(0.022)		(0.023)		(0.026)
Other situations		0.063***		0.014		0.011
		(0.022)		(0.024)		(0.026)
Kindergarten		-0.072***		-0.034***		-0.054***
between 3 and 5		(0.007)		(0.006)		(0.007)
Start school at 6		-0.220***		-0.188***		-0.195***
		(0.022)		(0.022)		(0.023)
Start school at 7 or more		-0.295***		-0.304***		-0.248***
		(0.026)		(0.032)		(0.033)
Observations	111,128	92,100	111,268	92,268	111,268	92,268

Notes: Dependent variables are the individual standardized grades in each of the three subjects.
Standard errors clustered at school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%
Base categories for dummies: male, student Spain, parent education - did not finish compulsory studies, parent occupation - blue-collar, lives with the mother and more than one sibling, kindergarten less than 3

although it goes from positive to negative. For General Knowledge, the bilingual program has a negative and significant effect over the score. This is the only exam related to a subject taught in English of those measured in CDI exam. Therefore it looks like the additional effort made to learn English by using it as a language of instruction in a subject other than English comes at the cost of lower performance in learning that subject.

To make a more intensive and flexible use of observable characteristics, we estimate the diff–in-diff regression by groups of students that have similar observable characteristics. In this way the performance of treated students is compared with the performance of students with the same observable characteristics in non treated periods and schools. Table 8 reports results by parental education for those students that were born in Spain, do not have any special educational needs, and are not older than 12 years old.[12] These represent more than two thirds of the population of students. In estimates not reported here for brevity, we use the parents' profession to form groups in addition to education variables, but the qualitative conclusion is the same. Other characteristics are included as covariates in the regression, since it is not possible to construct totally homogeneous groups. The estimates in this table are those of the parameter associated with the variable Bilingual school 2004/05 in CDI exam 2009/10, that is, the effect of the program we want to estimate. As with estimates with covariates in Table 7, we only find significant effects for General Knowledge. However, these estimates by groups have the following features:

[12] 11-12 years is the theoretical age that corresponds with sixth grade, which is the grade at which the CDI exam is taken (see subsection 3.1).

Table 8: Separate Diff-in-Diff regressions for observable groups of students: estimated treatment effect by group

Groups by parents education	**Mathematics**	**Reading**	**General Knowledge**	Proportion
University	-0.027 (0.096)	-0.117 (0.128)	-0.107 (0.134)	36.36%
Post-compulsory secondary	-0.083 (0.121)	-0.210 (0.136)	-0.259** (0.120)	19.11%
Compulsory education or less	-0.115 (0.081)	-0.062 (0.134)	-0.338** (0.154)	12.33%

Notes: Dependent variables are the individual standardized grades in each of the three subjects.
The sample used for these estimates are students of Spanish origin (i.e. non-immigrants), not older than 12 years and that do not have special education needs. They are divided by parents education in three groups. Proportion is the % that each group represents over the total sample of students (including those groups like students older than 12 years whose diff-in-diff estimates are not presented here.)
The following covariates were included in these regression though not reported: dummies for year of the exam and bilingual schools, sex, occupation of the parents, composition of the household in which the student lives and age at which the student started to go to school, preschool or daycare.
Standard errors clustered at school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

for Mathematics and General Knowledge the estimated effect is more negative for students whose parents have a lower level of education; for Mathematics all of them continue to be non-significant, but for General Knowledge there is not a significant effect for students whose parents have university education whereas it is significant for all the other students. Moreover, the difference between the effect for the university group and the effect for the compulsory education group is significantly different from zero at 10%. Surprisingly, for Reading there is no clear pattern. In any case the effect over reading is not significant for any of the groups.

4.2 Robustness checks

If the described changes in the population of treated students are only due to observable characteristics, then estimates of equation (3) are correctly identifying the average effect of the program. However, to check the robustness of these estimates, in this section we explore further the potential reasons that could lead to an endogenous change in the population of treated students, with respect to non-treated students. Even though the beginning of the program was not anticipated, the treatment lasted for six years until we observed our outcome variable and during that period the following movements of students may occur due to the program:

1. In any cohort of sixth grade students there is a proportion that had to repeat a grade as a consequence of failing to make sufficient progress. If a student starting primary education in 2003/04 were to repeat a grade in a bilingual school, he would

go from a non-bilingual education to a bilingual one. Most of the classmates of that child would have started school in 2004/05 and, therefore, they would have already participated in the bilingual program for some years. These repeaters may prefer, or may be recommended to move to a school that does not have the bilingual program in the grade they have to repeat. If this is the case, the treated cohort for which we observe our outcome variable may have a smaller proportion of these repeaters. One would expect this factor to improve the outcomes of the treated schools, and hence its removal would tend to strengthen our results.

2. As a consequence of the bilingual program there could be more students repeating a grade than in the previous cohort in the same school. We would not observe the outcome for these repeaters because they are not yet in the sixth grade.

3. Other endogenous movements can be related with the fact that some of the treated schools have vacancies. As mentioned in section 3.2 vacancies can be a reason for a school to apply for the program. Having treated schools with vacancies gives the opportunity to students with a good level of English, that otherwise might not have attended these schools, to apply for one of the vacancies once the program has started. Since the treatment we evaluate started six years before we measure the outcome, new students could have been coming for these reasons during five years.[13]

4. Finally some students that were in a bilingual school when the program was implemented might dislike the program and they could decide to change school at any point between the year of introduction of the program and the outcome we observe. We conjecture that once we have taken out repeaters from this cohort (whom we do not observe even if they stay in the same school as we have already mentioned) there is a very small proportion of students in this group. This is plausible because if they decide to move they cannot go to a highly demanded school, since at this stage they have all their vacancies filled. Nevertheless we do not have data to support our guess.

For those students in bilingual schools taking the exam in 2009/10 (i.e. the treated cohort) we know who was already at this school when they were five years old. For these students the implementation of the program was not known when deciding to enroll in this school. We can use this information to detect bilingual schools with a very large proportion of students in the treated cohort who stayed in the school since they were five years old. This will avoid the bias due to new students coming to the school when the program was already in place. We select the 8 bilingual schools that have a proportion of students that were not in that school at five years old smaller or equal than 16%. Table 9 presents estimates of equations (2) and (3) (i.e. Diff-in-diff estimates) using as treated group only

[13] This does not mean that all the newcomers will come because of this reason. Some movements of students would have occurred regardless of the program (for example due to migration) and we control for this by observing the same school before the program.

Table 9: Diff-in-Diff with and without covariates. Bilingual schools with more than 16% of the students coming to the school after being five years old are excluded.

	Mathematics		**Reading**		**General Knowledge**	
	No x	With x	No x	With x	No x	With x
Constant	0.002	4.536***	0.001	3.098***	0.001	3.421***
	(0.015)	(0.133)	(0.014)	(0.132)	(0.014)	(0.137)
Year 2010	-0.001	-0.073***	0.000	-0.084***	0.002	-0.072***
	(0.012)	(0.011)	(0.013)	(0.012)	(0.015)	(0.015)
Bilingual school 2004/05	-0.151	-0.077	-0.013	0.086	-0.050	0.050
	(0.128)	(0.086)	(0.220)	(0.198)	(0.150)	(0.119)
Bilingual school 2004/05	0.077	-0.017	0.028	-0.092	-0.155	-0.273*
in CDI exam 2010	(0.116)	(0.104)	(0.214)	(0.213)	(0.122)	(0.142)
Observations	109,654	90,892	109,793	91,059	109,793	91,059

Notes: Dependent variables are the individual standardized grades in each of the three subjects. Standard errors clustered at school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1% Though not reported, estimates with x include the same covariates as in Table 7.

those eight schools and excluding from the sample the other 17 bilingual schools. The results are similar to the results in Table 7 using the whole sample. The only difference is that the estimated effects tend to be higher here, including a less negative effect on General Knowledge. Furthermore, the same results are obtained when doing the Diff-in-diff using as treated students only those that were at the treated schools before the announcement and introduction of the program.

A different approach to the diff-in-diff is to find a control group of schools that is as close as possible to the treated schools. We have information about the schools that applied to the program and the criteria announced to choose schools, mentioned in section 2. In particular, among the 192 schools that applied, 64 schools had more than 60 points (out of 70) in those criteria. The 25 selected were all from this group with scores above 60. The other 38 schools that were not selected but are comparable in these criteria form a natural control group. By assuming that these are comparable groups, we do not have to use the diff-in-diff strategy and we can run a regression using only the 2009/10 results of the exam. To ensure an adequate comparison of the population of students in the treated and control groups we include as covariates the characteristics of the students we observe, and we also estimate by IV using as an instrument the indicator of having been at the same school when the student was five years old (i.e. having being assigned to treatment). Table 10 contains these two estimates. Both OLS and IV estimates imply the same qualitative conclusions as in the rest of the estimates presented: negative and significant effect on General Knowledge of being in the bilingual program and no effect significantly different from zero on mathematics and reading.

Table 10: OLS and IV with Schools that applied to became a bilingual schools and scored high in the selection criteria.

	Mathematics		Reading		General Knowledge	
	OLS	IV	OLS	IV	OLS	IV
Constant	4.020***	4.086***	4.288***	4.245***	3.143***	3.235***
	(0.739)	(0.739)	(0.857)	(0.849)	(0.826)	(0.811)
Bilingual school 2004/05	**-0.070**	**-0.123**	**-0.081**	**-0.046**	**-0.186***	**-0.261****
in CDI exam 2009/10	(0.082)	(0.093)	(0.056)	(0.060)	(0.098)	(0.110)
Female	-0.249***	-0.247***	-0.115***	-0.116***	-0.182***	-0.179***
	(0.039)	(0.038)	(0.037)	(0.037)	(0.044)	(0.044)
Student with special	-0.876***	-0.875***	-0.783***	-0.784***	-0.718***	-0.717***
educational needs	(0.078)	(0.077)	(0.103)	(0.101)	(0.117)	(0.116)
Student with disability	-1.204***	-1.206***	-1.214***	-1.213***	-0.937***	-0.940***
	(0.083)	(0.083)	(0.129)	(0.127)	(0.119)	(0.118)
Student's age	-0.340***	-0.344***	-0.345***	-0.343***	-0.267***	-0.271***
	(0.058)	(0.058)	(0.070)	(0.069)	(0.067)	(0.065)
Student Latin America	-0.251***	-0.264***	0.061	0.069	0.012	-0.005
	(0.082)	(0.081)	(0.073)	(0.072)	(0.085)	(0.085)
Student China	0.777**	0.774**	-0.031	-0.028	0.032	0.028
	(0.372)	(0.371)	(0.263)	(0.257)	(0.220)	(0.220)
Parent education -	0.242***	0.243***	0.279***	0.278***	0.232**	0.233**
University	(0.086)	(0.085)	(0.101)	(0.100)	(0.093)	(0.093)
Parent education -	0.080	0.081	0.210**	0.209**	0.143	0.145
Higher secondary	(0.075)	(0.075)	(0.099)	(0.098)	(0.093)	(0.094)
Parent education -	0.055	0.057	0.243**	0.241**	0.142	0.145
Vocational training	(0.102)	(0.102)	(0.116)	(0.114)	(0.107)	(0.107)
Parent education -	-0.096	-0.095	0.128	0.127	-0.010	-0.007
Lower secondary	(0.086)	(0.086)	(0.094)	(0.093)	(0.100)	(0.100)
Parent occupation -Busi.	0.189***	0.190***	0.063	0.062	0.117**	0.120**
minister, city hall	(0.049)	(0.048)	(0.052)	(0.051)	(0.052)	(0.051)
Parent occupation-	0.268***	0.268***	0.133***	0.133***	0.088*	0.088**
Professional	(0.051)	(0.050)	(0.050)	(0.049)	(0.045)	(0.044)
Start school at 6	-0.463***	-0.454***	-0.196	-0.202	-0.162	-0.149
	(0.150)	(0.152)	(0.205)	(0.202)	(0.200)	(0.202)
Start school	-0.405***	-0.410***	-0.003	-0.000	0.012	0.006
at 7 or more	(0.125)	(0.123)	(0.219)	(0.217)	(0.167)	(0.163)
Observations	2,177	2,177	2,192	2,192	2,192	2,192
R-squared	0.288	0.287	0.194	0.194	0.165	0.163

Notes: Dependent variables are the individual standardized grades in **2009/10** CDI exam in each of the three subjects.

Standard errors clustered at school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Reference categories for dummies and explanatory variables includes in the estimates are as in equations with covariates in Table 7. However, explanatory variables with no significant coefficient in any equation or those variables related with composition of the family living with the student are not reported here.

4.3 Estimates with an additional year of data

The estimates from sections 4.1 and 4.2 report the effect of the program on the first cohort of students treated in the group of 25 schools that first implemented the program. In 2009/10 this cohort finished sixth grade, the last year of primary education, and took the CDI exam. We have used that exam as measure of the outcome. Likewise, we can use the results of the sixth graders in the CDI exam in 2010/11 as the output for the second cohort of students treated at those 25 schools, and the output for the first cohort of students treated in the 54 schools selected in 2005/06 to implement the program.[14] The availability of this additional year of data allows us to test whether there are any improvements in the second cohort of treated students in the first 25 schools. It also allows us to check if our results for the schools selected in 2004 to participate are confirmed for the schools selected in 2005, since, as explained in Section 2, there were some significant changes in the selection criteria from one year to the next.

4.3.1 Results for the second cohort of students in the schools selected in 2004/05

Table 11: Diff-in-Diff with and without covariates. Second class of students treated at the 25 schools selected to implement the bilingual program in 2004/05. Comparing CDI 2011 with CDI 2009.

	Mathematics		**Reading**		**General Knowledge**	
	No x	With x	No x	With x	No x	With x
Constant	0.006	4.451***	0.007	2.859***	0.004	3.548***
	(0.015)	(0.140)	(0.014)	(0.124)	(0.015)	(0.132)
Year 2011	-0.004	-0.022*	-0.022*	0.067***	0.001	-0.016
	(0.014)	(0.013)	(0.013)	(0.011)	(0.014)	(0.014)
Bilingual school 2004/05	-0.049	0.041	0.041	0.020	-0.049	0.075
	(0.097)	(0.092)	(0.092)	(0.04)	(0.093)	(0.094)
Bilingual school 2004/05 in CDI exam 2010/11	0.022 (0.097)	-0.082 (0.096)	-0.082 (0.096)	-0.027 (0.048)	-0.076 (0.090)	-0.210*** (0.091)
Observations	110,939	91,681	110,966	91,705	110,966	91,705

Notes: Dependent variables are the individual standardized grades in each of the three subjects in 2009 and 2011. Standard errors clustered at school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Though not reported, estimates with x include the same covariates as in Table 7.

The descriptive statistics for the second cohort of students (cohort of 2010/11) being treated in the first 25 schools are very similar to those reported in Table 2 for the treated cohort of 2009/10 and they are not reported here to save space. Table 11 reports the

[14] Each new selected school starts implementing the program in the first grade and expands it to the other grades, year by year, until all the primary education classes in those schools follow the bilingual program.

estimated effect for this second treated cohort of students. The qualitative conclusion is the same as with the first cohort of treated students, presented and discussed in the previous two subsections. Quantitatively, the estimates tend to be greater (including a less negative effect on General Knowledge) than those reported in Table 7, but the differences are small. In any case, this small improvement in the second cohort is not enough to make the negative average effect on General Knowledge insignificant.

4.3.2 Results for the first cohort of students in the schools selected in 2005/06

The descriptive statistics for the first cohort of treated students in the 54 schools selected to implement the program in 2005/06 are in Table 12. The demographic characteristics of the last cohort of non-treated students at these schools are closer to the general population characteristics than those in the last non-treated cohort of the 25 schools. This can be seen by looking at the differences between the first two columns in Table 12 and comparing it with those differences in Table 2. Also, the change in demographic characteristics observed when comparing the last non-treated cohort with the first treated cohort is slightly smaller here than in the first 25 schools selected to implement the program.

Next, we look at the estimated effects of the treatment by observable groups of students for the 54 schools that became bilingual in 2005/06. These estimates are reported in Table 13. We see that, as in the previous analysis for the first 25 schools selected, the effect is not significantly different from zero in mathematics and reading. However, for General Knowledge the effect is now non-significant. This change in the average estimated effect could be due to a composition effect, since the effect is heterogeneous. As seen in Table 8 the effect is higher in absolute value the smaller the level of education of the parents. The students at these 54 school have better socio-demographic characteristics than those at the first 25 bilingual schools for which we detected a negative and significant effect in General Knowledge. This is why we next look at the estimated effects by groups of observables.

We can see in Table 14 that here the effects in mathematics and reading continue being not significant for any group. Also, as for the first 25 bilingual schools, in General Knowledge the effect is heterogeneous, and it is clearly non-significant for those students whose parents have a college degree, and negative and significant for those whose parents have only compulsory education or less. However, there is an important difference with respect to the estimated effect of the treatment in the first 25 schools presented in the previous sections. The negative effect of the program is smaller (in absolute value) here. This change implies that, for those students whose parents have post-compulsory secondary education the effect of the program in General Knowledge is now not significantly different from zero. The estimated effect is now -0.033 and in Table 8 it was -0.259.[15] Also, all the other estimates for the effect in General Knowledge (column 3 in Table 14) and most of the other estimates in this Table are much smaller (in absolute value) than the estimated effects for the first 25 schools.

[15] A test of equality of these two estimated effects rejects the null hypothesis of equality of effects.

Table 12: Descriptive statistics for the 2005/06 bilinguals schools

	Treat. bef.	**Cont. bef.**	**Treat. aft.**	**Cont. aft.**	Diff-in-Diff
Variable	Mean	Mean	Mean	Mean	
Subjects					
Dictation	7.61	7.90	3.54	3.70	
Mathematics	10.44	10.91	5.61	5.90	
Language	14.48	14.86	7.33	7.56	
Reading	3.54	3.59	3.80	3.87	
General knowledge	3.34	3.37	5.39	5.53	
Subjects - standard.					
Dictation	-0.10	0.00	-0.11	0.00	-0.01
Mathematics	-0.08	0.01	-0.09	0.00	0.00
Language	-0.08	0.00	-0.09	0.00	-0.01
Reading	-0.03	0.00	-0.05	0.00	-0.02
General knowledge	-0.02	0.00	-0.05	0.00	-0.03
Individual charac.					
Female	0.47	0.49	0.46	0.49	-0.01
Stud. with special ed.	0.07	0.06	0.08	0.06	0.01
Student with disab.	0.04	0.03	0.04	0.03	0.00
Student's age	12.17	12.14	12.13	12.15	-0.05
Student Spain	0.76	0.82	0.81	0.82	0.05
Student Romania	0.03	0.02	0.03	0.02	0.00
Student Morocco	0.01	0.01	0.01	0.01	0.00
Student Lat.Am.	0.13	0.10	0.09	0.09	-0.03
Student China	0.00	0.01	0.00	0.01	0.00
Student other	0.06	0.05	0.05	0.05	-0.01
Parent education					
Univ.	0.39	0.48	0.45	0.49	0.05
Higher secondary	0.21	0.18	0.20	0.18	-0.01
Vocational training	0.11	0.12	0.12	0.12	0.01
Lower secondary	0.22	0.17	0.18	0.16	-0.03
Did not finish comp.	0.07	0.05	0.05	0.05	-0.02
Parent profession					
Business, civil serv.	0.18	0.22	0.20	0.22	0.02
Professional	0.27	0.33	0.30	0.34	0.02
Blue Collar	0.55	0.44	0.50	0.44	-0.05
Age start. sch.					
Start school before 3	0.52	0.54	0.56	0.55	0.03
Pre-school 3-5	0.44	0.42	0.41	0.42	-0.03
Start school at	0.03	0.02	0.02	0.02	-0.01
Start sch. after 6	0.02	0.01	0.01	0.01	-0.02
Obs. Schools	54	1163	54	1179	
Obs. Students	2074	51076	2072	54807	

Table 13: Diff-in-Diff with and without covariates. First class of students treated at the 54 schools selected to implement the bilingual program in 2005/06.

	Mathematics		**Reading**		**General Knowledge**	
	No x	With x	No x	With x	No x	With x
Constant	0.005	5.175***	0.002	3.265***	0.004	3.718***
	(0.015)	(0.139)	(0.012)	(0.136)	(0.014)	(0.138)
Year 2011	-0.000	0.041***	0.000	0.067***	0.001	0.058***
	(0.012)	(0.012)	(0.011)	(0.011)	(0.014)	(0.014)
Bilingual school 2005/06	-0.084	0.005	-0.037	0.020	-0.025	0.069
	(0.065)	(0.050)	(0.056)	(0.040)	(0.074)	(0.064)
Bilingual school 2005/06	-0.014	-0.058	-0.014	-0.027	-0.031	-0.084
in CDI exam 2010/11	(0.063)	(0.058)	(0.049)	(0.048)	(0.069)	(0.066)
Observations	109919	95892	110029	96034	110029	96034

Notes: Dependent variables are the individual standardized grades in each of the three subjects in 2010 and 2011. Standard errors clustered at school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Though not reported, estimates with x include the same covariates as in Table 7.

Table 14: Diff-in-Diff for the 2005/06 schools. Estimated treatment effects using separate regressions by observable groups of students.

Groups by parents education	Mathematics	Reading	General Knowledge	Proportion
University	-0.101	-0.069	-0.017	37.53%
	(0.076)	(0.061)	(0.076)	
Post-compulsory secondary	-0.005 (0.074)	-0.014 (0.086)	-0.033 (0.085)	19.92%
Compulsory education or less	-0.058 (0.128)	-0.098 (0.067)	-0.196* (0.110)	11.76%

Notes: Dependent variables are the individual standardized grades in each of the three subjects in CDI exams in 2010 and 2011.

The sample used for these estimates are students of Spanish origin (i.e. non-immigrants), not older than 12 years and that do not have special education needs. They are divided by parents education in three groups. Proportion is the % that each group represents over the total sample of students (including those groups like students older than 12 years whose diff-in-diff estimates are not presented here.)

The following covariates were included in these regression though not reported: dummies for year of the exam and bilingual schools, sex, occupation of the parents, composition of the household in which the student lives and age at which the student started to go to school, preschool or daycare.

Standard errors clustered at school level in parentheses. * significant at 10%.

What can explain the different effects of the program found between the 25 schools selected to implement the program in 2004/05 and the 54 schools selected in 2005/06? Given that the characteristics of the students are different in these two groups of schools, the differential effect might be capturing positive peer effects in the 54 schools. To check this hypothesis we estimate our models including as explanatory variables the average parent's education levels of the students in each school. These variables are not significantly different from zero and the estimated effects of the policy do not change. Another explanation could be that those selected in 2005/06 are more suited and better prepared to implement the program so that the negative effect observed in the 25 schools is mitigated. As explained in section 2, in 2005/06 the English level of teachers in candidate schools was evaluated with an exam and the result in that exam was part of the criteria used to select schools. This may imply that the schools selected in 2005/06 were more prepared to teach in English. If this hypothesis is correct, it would imply that a great part of the negative effect found for the 25 bilingual schools from 2004/05 is due to an insufficient previous English training of the teachers in the schools selected. This is only a conjecture, which at this point we cannot test with the data available to us.

5 Concluding Remarks

All our estimates controlling for observable students' characteristics and using different ways for controlling self-selection in order to isolate the effect of the bilingual program on Mathematics, Reading, and General Knowledge lead to the same conclusion: there is a clear negative effect, and quantitatively substantial, on learning the subject taught in English, and no effect significantly different from zero on mathematical and reading (in Spanish) skills. The outcome variable used to measure learning in these three subjects is a general standardized exam on the basic skills that any student in sixth grade is supposed to have acquired during the primary school years.

Two aspects of the results are particularly important because of their potential policy implications. The first one is that the negative effects are concentrated on the children of less educated parents. The second one is that the negative effect is much larger (in absolute value) for the group of schools that started participating in 2004 than for those that started in 2005. This even makes the negative effect not significantly different from zero on average and for the students whose parent have more than lower-secondary education. From 2004 to 2005 there was a change in the rules that increased the required English knowledge of the teachers at participating schools. It would be worth ascertaining to which extent this is the cause of the decrease in the negative impact.

Given the change in observable characteristics of the students after the introduction of the program, a change in unobservable characteristics might be suspected. This might bias our estimates. Given the different sources of the change in the population of students in bilingual schools, the direction of the bias is uncertain. However, it is not unreasonable to assume that the change in unobservable characteristics is the same as in the observable ones. If that were the case, this would reinforce our negative and significant effect on General Knowledge and it might turn the estimated insignificant effect on mathematics and reading into a negative and significant effect. On the other hand, if observables and unobservables are positively correlated, the observable characteristics should already be picking up much of the effect of unobservables and for this reason the effect of the program would not differ much from our current estimates, especially if the positive correlation between observables and unobservables is very high. The difficulties we experience in being certain about the effects of the policy is a stark reminder about the necessity of introducing policies in a way that makes it possible its correct evaluation. This is particularly unforgivable in a context like the present one, when the policy was introduced gradually and the applicants were all quite similar.

This study is based only on the first two cohorts of students finishing primary education in the bilingual program. The addition of more cohorts and more schools in future years may allow for a more detailed analysis. One particularly worthwhile aspect for further research is the reaction of parents when choosing schools once it is known at the time of entering preschool that the school is part of the bilingual program. We might observe a marked segregation of students. This will be specially strong in secondary education, when having performed well in the bilingual program is a requirement to enroll in bilingual sections of High schools. The long run effect of the program, and the potential segregation are important avenues for further research.

Finally, as mentioned in the Introduction, the fact that Admiraal, Westhoff and de Bot (2006) found no effect of a similar program on secondary education students in The Netherlands opens the additional question of what is the best age for introducing a program like this.

References

[1] Admiraal, Wilfried, Gerard Westhoff and Kees de Bot (2006), Evaluation of Bilingual Secondary Education in The Netherlands: Students' language proficiency in English," *Educational Research and Evaluation* 12, 75 – 93.

[2] Angrist, Joshua D. and Victor Lavy (1999), "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics* 114, 533-575.

[3] Banfi, Cristina and Raymond Day (2004), "The Evolution of Bilingual Schools in Argentina," *International Journal of Bilingual Education and Bilingualism* 7, 398-411.

[4] Bentolila, Samuel and Andrea Ichino (2008), "Unemployment and consumption near and far away from the Mediterranean," *Journal of Population Economics* 21, 255–280.

[5] Bergin, Adele, Thomas Conefrey, John Fitz Gerald and Ide Kearney (2009),"Recovery Scenarios for Ireland: An Update," mimeo.

[6] Fidrmuc, Jarko and Jan Fidrmuc (2009), "Foreign Languages and Trade: What are you sinking about?," mimeo.

[7] Fry, Richard and Lindsay B. Lowell (2003), "The Value of Bilingualism in the U.S. Labor Market," *Industrial & Labor Relations Review* 57, 128-140.

[8] Ginsburgh, Victor A. and Juan Prieto-Rodriguez (2011) "Returns to Foreign Languages of Native Workers in the EU," *Industrial & Labor Relations Review* 64, 599-618.

[9] Greene, Jay A. (1998), "A Meta-Analysis of the Effectiveness of Bilingual Education," mimeo, University of Texas at Austin.

[10] Krueger, Alan B. (1999), "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics Volume* 114, 497-532.

[11] Jepsen, Christopher (2009), "Bilingual Education and English Proficiency," mimeo. University of Kentucky.

[12] Ordóñez, Claudia L. (2004), "EFL and Native Spanish in Elite Bilingual Schools in Colombia: A First Look at Bilingual Adolescent Frog Stories," *International Journal of Bilingual Education and Bilingualism* 7, 449-474.

[13] Saiz, Albert and Elena Zoido (2005), "Listening to What the World Says: Bilingualism and Earnings in the United States," *The Review of Economics and Statistics* 87, 523-538.

[14] Williams, Donald R. (2011), "Multiple Language Usage and Earnings in Western Europe," *International Journal of Manpower* 32, 372-393.

[15] Willig, Ann C. (1985), "A Meta-Analysis of Selected Studies on the Effectiveness of Bilingual Education," *Review of Educational Research* 55, 269-317.