EVALUATING A BILINGUAL EDUCATION PROGRAM IN SPAIN: THE IMPACT BEYOND FOREIGN LANGUAGE LEARNING

BRINDUSA ANGHEL, ANTONIO CABRALES and JESUS M. CARRO*

Bilingual education programs, which consist of doing a substantial part of the instruction in a language different from the native language of the students, exist in several countries like the United States, India, and Spain. While the economic benefits of knowing a second language are well established, the potential effects over the learning of other subjects have received much less attention. We evaluate a program that introduced bilingual education (in English and Spanish) in primary education in a group of public schools of the Madrid region in 2004. Under this program, students not only study English as a foreign language but also some other subjects (at least Science, History, and Geography) are taught in English. In order to evaluate the program, a standardized test for all sixth grade students in Madrid on the skills considered "indispensable" at that age is our measure of the outcome of primary education. Our results indicate that there is a clearly negative effect on the exam results for the subject taught in English, for children whose parents have less than upper secondary education. This negative effect is a composite of two phenomena: the effect of the program on the student's knowledge of the subject and a reflection of the student ability to do the test in their native language when English is the medium of instruction. Although we are not able to separate quantitatively these two effects, the composite effect has a relevant interest, because the results for exams taken in Spanish are the measures that determine academic progression in the Spanish system. In contrast with the previous result, there is no significant effect for anyone on mathematical and reading skills, which were taught in Spanish. (*JEL* H40, I21, I28)

I. 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% in Austria and 39% 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,

*We gratefully acknowledge the help of José Carlos Gibaja and Ismael Sanz to obtain the data, the comments from Gema Zamarro and seminar participants at various conferences, and the support from the Spanish Ministry of Science and Technology from grants ECO2009-07530 and ECO2012-31985 (B.A.), ECO2012-34581 (A.C.), ECO2009-11165 and ECO2012-31358 (J.M.C.), CONSOLIDER-INGENIO 2010—CSD2006-0016 (all), and Consejera de Educación de la Comunidad de Madrid under the Excelecon project.

Anghel: Assistant Professor, Dpto. Análisis Económico, Universidad Autónoma de Madrid and FEDEA, Madrid, 28049, Spain. Phone +34 91497 7040, Fax +34 91577 9575, E-mail banghel@fedea.es

Cabrales: Professor, Department of Economics, University College London, London, WC1H 0AX, UK. Phone +44 203 108 5229, Fax +44 207 916 2775, E-mail a.cabrales@ucl.ac.uk

Carro: Associate Professor, Department of Economics, Universidad Carlos III de Madrid, Getafe, 28903, Spain. Phone +34 91624 9586, Fax +34 91624 9875, E-mail jcarro@eco.uc3m.es

1. An earlier analysis of the same data by Williams (2011) found a smaller impact: between 5% in Austria and 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

ABBREVIATIONS

CDI: prueba de Conocimientos y Destrezas Indispensables
diff-in-diff: Difference-In-Differences
IV: Instrumental Variable
OLS: Ordinary Least Squares

the country as a whole may also benefit: Fidrmuc and Fidrmuc (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 Ordóñez (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 in the academic year 2013–2014 has 406 public schools (316 primary schools and 90 high schools, around 40% of the total) where around 40% of the instruction, including all the science curriculum, is taught in English.[3] These programs have been so successful with voters that both major parties included in their 2011 general election platforms the promise of extending the program to the whole nation.[4]

This expansion of bilingual programs where at least part of the instruction is in a foreign language (i.e., different from the mother tongue of students) is certainly not a Spanish phenomenon. Other important examples are the English schools in India (Munshi and Rosenzweig 2006) and the one-way foreign language immersion programs for native English speakers in the United States (Center for Applied Linguistics 2011).

on wages, or Saiz and Zoido (2005) who find an effect of about 5%.

3. Andalusia also has a bilingual program, but the percentage of instruction in English is smaller, around 20% of the instruction time.

4. See, for example, 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" (http://www.psoe.es/saladeprensa/docs/608866/page/programa-electoral-para-las-elecciones-generales-2011.html) or the one of conservative party PP, which states "we will promote Spanish–English bilingualism in the whole educational system from pre-school to university" (http://www.pp.es/actualidad-noticia/programa-electoral-pp&uscore;5741.html).

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 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 that have received much less attention. The aim of this study 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 of 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 sixth grade students (12–13 years of age), starting with the school year 2004/2005. The exam tests for what are called "Indispensable Knowledge and Skills" (CDI in its Spanish acronym) 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 2004/2005, and we checked the results of the first- and second-treated student cohorts, which took the exam in 2009/2010 and in 2010/2011, respectively. 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 2010/2011.

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 sixth grade. To control for student 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. That is, in order to control for endogeneity problems, we use a *difference-in-difference* (diff-in-diff) approach with controls, comparing the exam results of children in the treated schools before and after they became bilingual with the group of nonbilingual schools before and after the treatment. Other ways of controlling for endogeneity, like using as instrument being enrolled at treated schools before the announcement of the program which is a proxy for exogenous "assignment to treatment" of students, confirm the diff-in-diff estimates.

For the first-treated cohort, 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.[5] As General Knowledge is the only subject taught in English from the three present in the exam, it seems clear that the extra effort made to use English as the medium of instruction comes at the expense of a worsening in the results of standard examinations of that subject in Spanish. In a sense, there is a confound because it is possible that the students do not know less, but simply they do not know how to express it in Spanish.[6] But, even if that is the case, this would also suggest that the level of linguistic competence in English is not enough to leap through that barrier. And, possibly more importantly, other standardized examinations which, unlike the CDI, do have academic consequences (at the end of secondary and at the entry to university) are in Spanish, so a negative result in CDI is still an outcome of interest.

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 because of a better selection of those schools in terms of the English knowledge of the teachers, because 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 United States. This literature finds mostly positive results of those programs. Willig (1985) concludes that the better the experimental design of the study, the more positive results 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 the foreign language programs aimed to immerse native English speakers in a foreign language in the United States, or regarding bilingual education in English for countries whose official language is not English. An exception

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

6. Hofstetter (2003) uses data of middle-school students in a low-income, predominately Latino area of southern California to study the effect for English language learners of different linguistic accommodations of standardized tests in Mathematics. What is relevant for our study is that part of the sample considered by Hofstetter (2003) was instructed in English. Of those students, some were tested in the language of instruction (English), while others were tested in their native language (Spanish). Hofstetter (2003) finds a negative effect of this particular mismatch between the language of instruction and the language of the test. Unfortunately, the institutional and the socioeconomic framework for these data are very different from the ones in our study, so we think that it would not be prudent to draw quantitative conclusions from Hofstetter (2003) about which part of the effect we find in general knowledge is due to the mismatch between instruction and testing language.

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, because 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.

This study is organized as follows. Section II describes the institutional setup and the program in some detail. Section III discusses the data and the econometric model. Section IV contains the main results of this study and it has some additional estimations and robustness checks. Section V contains conclusions.

II. 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 26% and a youth unemployment rate of 57%, only 39,690 Spaniards emigrated in the first semester of 2013. This contrasts markedly with the over 6 million unemployed, or with the 40,000 yearly emigrants that Bergin et al. (2009) estimate for Ireland, a country ten 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 (educacion secundaria obligatoria) and 2 years of noncompulsory education, which is divided into vocational training (ciclos formativos) and preparation for college (bachillerato). There are also 3 years of free publicly funded preschool, from ages 3 to 5. More than 96% of the students in the Madrid Region attended preschool. The preschool children share the premises with those in primary school. Also, the preschoolers 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 3 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 3 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 preschool. Since at the time the bilingual program was designed and announced, there were students already in the preschool level at the selected schools, their parents' school choices were made 3 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 2004/2005 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. The students from the treated schools from cohorts starting before the first year of the treatment do not participate in the program at all, and they remain with the same course of study as students in untreated schools. 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 cover only the schools from the first cohort. Once the students from the 2004/2005 cohort reached secondary education (in 2010/2011), a second phase kicked in, and some high schools joined the program. As 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 316 schools because of the high demand (out of a total of about 780 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 3 and 16.

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 five weekly periods (of 45–60 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 in both bilingual and nonbilingual schools[8] so that the margin of autonomy in the number of teaching hours in bilingual schools becomes clear.

With Knowledge of the Environment (a subject encompassing Science, Geography, and History) plus five periods of English, the minimum is accomplished (remember that a school is considered bilingual if instruction is carried out in English *at least* one third of the school time). 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). Expansion of English instruction from the minimum depends on the availability of teachers, but most schools end up having above 40% 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, because of 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 3 hours per week for teachers, and 4 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

7. In fact, there were 26 schools that became bilingual in 2004/2005, out of which we have enough information on 25 schools.

8. The schedule in Table 1 applies to all schools in each category (bilingual and nonbilingual). In Spain, schools have no freedom in the number of hours the teachers dedicate to each of the subjects.

TABLE 1

Weekly Schedule by Area in Primary School, NonBilingual and Bilingual Schools

	Number of Weekly Hours			Number of Weekly Hours		
	NonBilingual Schools			Bilingual Schools		
Areas	First Cycle First and Second Grades	Second Cycle Third and Fourth Grades	Third Cycle Fifth and Sixth Grades	First Cycle First and Second Grades	Second Cycle Third and Fourth Grades	Third Cycle Fifth and Sixth Grades
Knowledge of the environment	4	4	4	2.5	2.5	2.5
Art	3	2.5	2.5	1.5	1.5	1.5
Physical education	3	3	2.5	2	1.5	1.5
Spanish language	5	5	5	5	5	5
Foreign language	2	2.5	3	5	5	5
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
Extra hours				1	1.5	1.5
Total	25	25	25	25	25	25

Note: Extra hours in bilingual schools can be assigned to any English-taught subject, usually Knowledge of the Environment.

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 United Kingdom 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 into 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 it involves a 20-minute conversation with the examiner.

III. DESCRIPTION OF DATA AND ECONOMETRIC MODEL

A. *Description of Data*

Our data come from a standardized exam that has been administered each year in all primary schools from the Spanish region of Madrid to sixth grade students (12–13 years of age), starting with the school year 2004/2005.[9] 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 Organization for Economic Co-operation and Development's PISA exam, the CDI exam does not have any academic consequences for the students, 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 the exam scores as a measure of student achievement, 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 is filled out by each student.[10] 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,

9. Since the school year 2009/2010, the exam is also administered to all students in the third grade of compulsory secondary education (14–15 years old).

10. The exam results are anonymized and only the students and their families know individual results. For this reason, we cannot use any measure of lagged achievement in the analysis. Hence, for identification we need to rely on our diff-in-diff strategy and on the individual sociodemographic controls from the student questionnaire. If this identification strategy is valid, then the availability of information on lagged achievement would only serve to gain more precision.

Romania, and other, to have a sufficient number of observations in 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 or preschool. From the exam, we have information at student level on gender, whether the student has any special educational needs and 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: because 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 (e.g., teacher, researcher, doctor, engineer, lawyer, psychologist, artist, etc.); business and administrative occupations (e.g., chief executive officer, civil servant, etc.); and blue-collar occupations (e.g., shop assistant, fireman, construction worker, cleaning staff, etc.).[11]

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.

First Group of Schools Implementing the Program. The dataset with more information available for our empirical analysis comes from the first cohort of treated students in bilingual schools in the region of Madrid. They started first grade of primary school in 2004/2005, and took the CDI exam in 2009/2010. This first-treated cohort is from the 25 schools that firstly implemented the bilingual program.[12]

In order to control for the endogeneity problems caused by self-selection of students and schools which we will explain below, we use a diff-in-diff approach. We compare the performance of children in the treated schools before and after they became bilingual with the group of nonbilingual schools before and after the treatment. Thus, we employ the data for 2008/2009 and 2009/2010 cohorts. The four groups that we analyze are the following: the group of bilingual schools in 2008/2009 (the treatment group before the treatment), the group of nonbilingual schools in 2008/2009 (the control group before the treatment), the group of bilingual schools in 2009/2010 (the treatment group after the treatment), and the group of nonbilingual schools in 2009/2010 (the control group after the treatment).

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, in the treated school, there is an increase in the proportion of Spanish students from 81% in the year before treatment to 87% in the first-treated cohort, which translates into 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 2014]). 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

11. Robustness checks using separately the education of each parent yield very similar results.

12. The schools first selected to implement the program in 2004/2005 were actually 26, but due to unknown reasons we do not observe one of those schools in the year before the first-treated students finish. Therefore, we have to restrict our analysis to the 25 schools for which we have information.

TABLE 2

Descriptive Statistics Benchmark

Variable	Treatment Before Mean	Control Before Mean	Treatment After Mean	Control After Mean	Diff-in-Diff	Standard Error (Diff-in-Diff)
Subjects						
Mathematics	8.94	9.54	10.55	10.88	0.26	0.230
Reading	2.87	2.93	3.53	3.59	0.01	0.062
General knowledge	2.28	2.35	3.17	3.37	−0.13	0.057
Subjects—standard						
Mathematics	−0.11	0.00	−0.06	0.00	0.05	0.042
Reading	−0.04	0.00	−0.04	0.00	0.00	0.042
General knowledge	−0.05	0.00	−0.15	0.00	−0.11	0.042
Individual characteristics						
Female	0.50	0.49	0.51	0.49	0.01	0.022
Student with special education	0.11	0.07	0.06	0.06	−0.04	0.011
Student with disability	0.04	0.03	0.03	0.03	−0.01	0.007
Student's age	12.15	12.14	12.12	12.14	−0.04	0.017
Student Spain	0.81	0.81	0.87	0.81	0.06	0.017
Student Romania	0.03	0.02	0.02	0.02	−0.01	0.007
Student Morocco	0.01	0.01	0.00	0.01	0.00	0.005
Student Latin America	0.10	0.11	0.06	0.10	−0.03	0.013
Student China	0.00	0.01	0.00	0.01	0.00	0.003
Student other	0.05	0.04	0.04	0.05	−0.01	0.009
Parent education						
University	0.33	0.48	0.39	0.47	0.07	0.023
Higher secondary	0.21	0.17	0.20	0.18	−0.02	0.018
Vocational training	0.12	0.12	0.14	0.12	0.01	0.015
Lower secondary	0.26	0.17	0.22	0.17	−0.04	0.017
Did not finish compulsory	0.08	0.06	0.05	0.05	−0.02	0.011
Parent profession						
Business, civil servant	0.17	0.22	0.19	0.22	0.02	0.018
Professional	0.24	0.33	0.29	0.33	0.05	0.020
Blue collar	0.58	0.46	0.51	0.45	−0.06	0.022
Age starting school						
Start school before 3	0.46	0.51	0.51	0.54	0.02	0.022
Preschool 3–5	0.49	0.44	0.47	0.43	0.00	0.021
Start school at 6	0.03	0.03	0.02	0.02	−0.01	0.007
Start school after 6	0.02	0.01	0.01	0.01	−0.01	0.005
Observation of schools	25	1,201	25	1,217		
Observation of students	1,135	55,793	1,145	53,150		

both years (at most there is a difference of one decimal).

The numbers presented above indicate that there has been a change (certainly not large and perhaps endogenous) in the characteristics of the students enrolled in the bilingual schools from the period before to the one after the treatment. This change involves an improvement in student characteristics, such as the level of education and the occupation of parents, or their nationality, which are known to be determinants of the academic performance of children. The same change could be taking place in other unobservable determinants. In our analysis of this issue, we will use the additional information that we describe in the following paragraphs to account for these changes.

We obtained the list of children who attended the treated schools since they were 5 years old, the last year of preschool education. With that list, first, we analyze the group of schools where the number of children who entered after they became bilingual (i.e., children who were not enrolled in that school when they were 5 years old) is less than 4 (i.e., about 16% in the average class of 25). There are eight treated schools that satisfy this condition. As before, we compare these schools before they became bilingual (the 2008/2009 cohort) and after they became bilingual (the 2009/2010 cohort) and we use as a control group the group of nonbilingual 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

TABLE 3
Descriptive Statistics—Schools with Few Movements

Variable	Eight School Before Mean	Eight School After Mean
Subjects		
Mathematics	8.73	10.48
Reading	2.92	3.61
General knowledge	2.28	3.11
Subjects—standard		
Mathematics	−0.15	−0.07
Reading	−0.01	0.02
General knowledge	−0.05	−0.20
Individual characteristics		
Female	0.49	0.50
Student with special education	0.08	0.07
Student with disability	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 Latin America	0.10	0.05
Student China	0.00	0.00
Student other	0.03	0.01
Parent education		
University	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 compulsory	0.08	0.05
Parent profession		
Business, civil servant	0.17	0.20
Professional	0.23	0.26
Blue collar	0.60	0.54
Age starting school		
Start school before 3	0.46	0.55
Preschool 3–5	0.52	0.44
Start school at 6	0.02	0.01
Start school after 6	0.01	0.00
Observation of schools	8	8
Observation of students	416	434

observe an important increase in the proportion of students whose parents have university degrees, from 27% in the pretreatment cohort to 36% in the posttreatment 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 preschool before 3 years old increases by 6% 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 since they were 5 years of age. The introduction of the bilingual education program was not announced in advance of enrolling those children in the treated schools. 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 2004/2005, 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 socioeconomic 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

TABLE 4
Descriptive Statistics—Children Who Did Not Move

Variable	Treatment Before Mean	Control Before Mean	Treatment After Mean	Control After Mean	Diff-in-Diff	Standard Error (Diff-in-Diff)
Subjects						
Mathematics	8.94	9.54	10.54	10.88	0.25	0.249
Reading	2.87	2.93	3.57	3.59	0.05	0.067
General knowledge	2.28	2.35	3.16	3.37	−0.14	0.062
Subjects—standard						
Mathematics	−0.11	0.00	−0.06	0.00	0.05	0.046
Reading	−0.04	0.00	−0.01	0.00	0.03	0.046
General knowledge	−0.05	0.00	−0.16	0.00	−0.12	0.046
Individual characteristics						
Female	0.50	0.49	0.51	0.51	−0.02	0.023
Student with special education	0.11	0.07	0.05	0.07	−0.06	0.011
Student with disability	0.04	0.03	0.04	0.03	−0.01	0.007
Student’s age	12.15	12.14	12.09	12.15	−0.07	0.018
Student Spain	0.81	0.81	0.93	0.81	0.11	0.018
Student Romania	0.03	0.02	0.01	0.02	−0.02	0.007
Student Morocco	0.01	0.01	0.01	0.01	0.00	0.005
Student Latin America	0.10	0.11	0.04	0.10	−0.06	0.014
Student China	0.00	0.01	0.00	0.01	0.00	0.003
Student other	0.05	0.04	0.02	0.05	−0.03	0.010
Parent education						
University	0.33	0.48	0.38	0.47	0.05	0.025
Higher secondary	0.21	0.17	0.20	0.18	−0.02	0.019
Vocational training	0.12	0.12	0.14	0.12	0.01	0.016
Lower secondary	0.26	0.17	0.24	0.17	−0.02	0.019
Did not finish compulsory	0.08	0.06	0.05	0.05	−0.02	0.011
Parent profession						
Business, civil servant	0.17	0.22	0.20	0.22	0.02	0.019
Professional	0.24	0.33	0.27	0.33	0.02	0.022
Blue collar	0.58	0.46	0.53	0.45	−0.05	0.023
Age starting school						
Start school before 3	0.46	0.51	0.52	0.54	0.03	0.023
Preschool 3–5	0.49	0.44	0.47	0.43	0.00	0.023
Start school at 6	0.03	0.03	0.01	0.02	−0.02	0.007
Start school after 6	0.02	0.01	0.00	0.01	−0.01	0.005
Observation of schools	25	1,201	25	1,217		
Observation of students	1,135	55,793	849	53,150		

is bigger than in the bilingual schools. Second, the characteristics of children change from the pretreatment cohort to the posttreatment 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 2008/2009 before the treatment. However, in the posttreatment, 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 in posttreatment period.

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.

Second cohort of students in the first schools implementing the program. We have data for the second cohort of students (class of 2010/2011) being treated in the first 25 schools implementing the program. They started primary school in the year 2005/2006. They are 1 year younger than the first cohort of treated students but they too were already enrolled as preschool students when the program was announced. The descriptive statistics for them are very similar to those reported in Table 2 for the treated cohort after the treatment and they are not reported here to save space. We will estimate the effect for this second-treated cohort to see if there is any learning in these schools from having implemented the program to the first cohort of students.

TABLE 5
Descriptive Statistics—Children Who Moved

Variable	Mean	Standard Deviation
Subjects		
Mathematics	10.62	5.86
Reading	3.42	1.50
General knowledge	3.26	1.24
Subjects—standard		
Mathematics	−0.05	1.07
Reading	−0.11	1.05
General knowledge	−0.08	0.98
Individual characteristics		
Female	0.49	0.50
Student with special education	0.12	0.33
Student with disability	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 Latin America	0.12	0.33
Student China	0.01	0.09
Student other	0.10	0.30
Parent education		
University	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 compulsory	0.06	0.25
Parent profession		
Business, civil servant	0.20	0.40
Professional	0.35	0.48
Blue collar	0.45	0.50
Age starting school		
Start school before 3	0.47	0.50
Preschool 3–5	0.46	0.50
Start school at 6	0.05	0.22
Start school after 6	0.02	0.14
Observation of schools	26	
Observation of students	341	

TABLE 6
Descriptive Statistics—Schools That Applied to Become a Bilingual School and Scored High in the Selection Criteria

Variable	Mean in Pretreatment Period	Mean in Posttreatment Period
Subjects		
Mathematics	8.32	10.31
Reading	2.46	3.51
General knowledge	2.06	3.34
Subjects—standard		
Mathematics	−0.22	−0.10
Reading	−0.32	−0.06
General knowledge	−0.20	−0.02
Individual characteristics		
Female	0.47	0.47
Student with special education	0.09	0.09
Student with disability	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 Latin America	0.17	0.16
Student China	0.00	0.01
Student other	0.06	0.06
Parent education		
University	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 compulsory	0.10	0.07
Parent profession		
Business, civil servant	0.19	0.17
Professional	0.22	0.27
Blue collar	0.59	0.56
Age starting school		
Start school before 3	0.46	0.52
Preschool 3–5	0.49	0.44
Start school at 6	0.03	0.02
Start school after 6	0.02	0.02
Observation of schools	38	38
Observation of students	1, 341	1, 292

Second Group of Schools Implementing the Program. A second group of 54 schools were selected to implement the program from 2005 to 2006. These were added to the 25 schools that started implementing the program in 2004/2005. We have data for the first cohort of treated students in these 54 schools. They finished primary education and took the CDI exam in 2010/2011. We analyze the results for these treated students separately from the students from the first 25 schools implementing the program for two reasons. First, there were some changes in the criteria used to select schools, as explained in Section II. Second, the class of 2010/2011 from the 25 schools is the second cohort treated at those schools, whereas these are the first cohort treated at the 54 schools.

Only 53 of the 54 schools are going to be used in our study. One school is considered to be an outlier because at the same time it has a very large (the tenth largest among 1,226 schools) increase on the average reading score from 2009 to 2010, and a very large (the fourth largest among 1,226 schools) reduction on the average reading score the next year. Furthermore, such large and contradictory changes only happen in Reading; they do not happen in Mathematics or General Knowledge. Given this, we decided to exclude this school from our analysis.[13]

The descriptive statistics for the first cohort of treated students in the 53 schools selected

13. Our estimates of the effect of the program were made including and excluding this observation and there is almost no change.

TABLE 7
Descriptive Statistics for the 2005/2006 Bilingual Schools

Variable	Treatment Before Mean	Control Before Mean	Treatment After Mean	Control After Mean	Diff-in-Diff	Standard Error (Diff-in-Diff)
Subjects						
Mathematics	10.42	10.91	5.61	5.90	0.20	0.139
Reading	3.53	3.59	3.80	3.87	−0.01	0.044
General knowledge	3.34	3.37	5.39	5.53	−0.11	0.064
Subjects—standard						
Mathematics	−0.08	0.01	−0.09	0.00	0.00	0.032
Reading	−0.04	0.00	−0.04	0.00	−0.02	0.032
General knowledge	−0.02	0.00	−0.05	0.00	−0.032	
Individual characteristics						
Female	0.47	0.49	0.46	0.49	−0.01	0.016
Student with special education	0.07	0.06	0.08	0.06	0.01	0.008
Student with disability	0.04	0.03	0.04	0.03	0.00	0.005
Student's age	12.17	12.14	12.13	12.15	−0.05	0.012
Student Spain	0.76	0.82	0.81	0.82	0.05	0.012
Student Romania	0.03	0.02	0.03	0.02	0.00	0.005
Student Morocco	0.01	0.01	0.01	0.01	0.00	0.003
Student Latin America	0.13	0.10	0.09	0.09	−0.03	0.009
Student China	0.01	0.01	0.00	0.01	0.00	0.002
Student other	0.06	0.05	0.05	0.05	−0.01	0.007
Parent education						
University	0.39	0.48	0.45	0.49	0.05	0.017
Higher secondary	0.21	0.18	0.20	0.18	−0.01	0.013
Vocational training	0.11	0.12	0.12	0.12	0.01	0.011
Lower secondary	0.22	0.17	0.18	0.16	−0.03	0.012
Did not finish compulsory	0.07	0.05	0.05	0.05	−0.02	0.007
Parent profession						
Business, civil servant	0.18	0.22	0.20	0.22	0.02	0.013
Professional	0.27	0.33	0.30	0.34	0.02	0.015
Blue collar	0.55	0.44	0.50	0.44	−0.05	0.016
Age starting school						
Start school before 3	0.51	0.54	0.56	0.55	0.03	0.016
Preschool 3–5	0.44	0.42	0.41	0.42	−0.03	0.016
Start school at 6	0.03	0.02	0.02	0.02	−0.01	0.005
Start school after 6	0.02	0.01	0.01	0.01	−0.02	0.003
Observations of schools	53	1,163	53	1,179		
Observations of students	2,057	51,076	2,056	54,807		

to implement the program in 2005/2006 are reported in Table 7. The demographic characteristics of the last cohort of nontreated students at these schools are closer to the general population characteristics than those in the last nontreated cohort of the 25 schools. This can be seen by looking at the differences between the first two columns in Table 7 and comparing it with those differences in Table 2. Moreover, the change in demographic characteristics observed when comparing the last nontreated cohort with the first-treated cohort is smaller here than in the first 25 schools selected to implement the program.

B. Econometric Model of Education Production

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

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

where x_i is the observable characteristics of students and their families are described in Section III.A, bil_j indicates whether school j participated in the bilingual program, u_i is unobservable characteristics of the students, such as effort or ability, v_j is 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 the 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 because of low quality) with teachers about to be displaced.[14]

Estimation Strategy. To control for the endogeneity problem caused by the self-selection of schools and students explained, we use diff-in-diff estimation. This solves the self-selection of schools into the program because we observe the same school the first year 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 mentioned in Section II since the admission rules to primary school gives precedence to preschoolers 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 are not expected to 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 nontreated 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 III.A, 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 these changes.

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

$$(2)\quad y_{ij} = \alpha_0 + \alpha_1 bil_j + \alpha_2 \,\text{year}\,2010 + \delta\left(\text{year 2010} \cdot bil_j\right) + \varepsilon_{ij}$$

$$(3)\quad y_{ij} = \alpha_0 + \alpha_1 bil_j + \alpha_2 \,\text{year}\;\;2010 + \delta\left(\text{year 2010} \cdot bil_j\right) + \beta x_i + \varepsilon_{ij}$$

where year 2010 is a dummy variable for the academic year 2009/2010, 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.

IV. RESULTS

A. *Estimates of the Effect of the Program for the First-Treated Cohort*

In Table 8, we present estimates of models (2) and (3). The parameter associated with the variable Bilingual school 2004/2005 in posttreatment period ($y10 \cdot bil_j$) 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, although it goes from positive to negative. For General Knowledge, the bilingual program has a negative and significant effect over the score. As General Knowledge is the only subject taught in English from the three present in the exam, it seems clear that the extra effort

14. 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 make significant efforts to avoid school closures.

TABLE 8

Diff-in-Diff with and without Covariates: All Students (Bilingual Schools 2004/2005)

	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)
Posttreatment period	−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 2004/2005	−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 2004/2005	0.053	−0.068	0.002	−0.110	−0.096	−0.229**
In posttreatment period	(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 educational needs		−0.744***		−0.702***		−0.620***
		(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—university		0.340***		0.273***		0.249***
		(0.016)		(0.018)		(0.018)
Parent education—higher secondary		0.182***		0.173***		0.169***
		(0.015)		(0.018)		(0.017)
Parent education—vocational training		0.181***		0.204***		0.184***
		(0.016)		(0.019)		(0.018)
Parent education—lower secondary		0.100***		0.105***		0.102***
		(0.015)		(0.019)		(0.017)
Parent occupation—business, minister, city hall		0.167***		0.139***		0.102***
		(0.010)		(0.010)		(0.011)
Parent occupation—professional		0.251***		0.205***		0.151***
		(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 and one sibling		0.071***		0.034		0.030
		(0.025)		(0.025)		(0.029)
Lives with both parents		0.066***		0.003		0.065**
		(0.022)		(0.023)		(0.026)
Lives with both parents and one sibling		0.174***		0.068***		0.100***
		(0.022)		(0.022)		(0.025)
Lives with both parents and more than one sibling		0.151***		0.055**		0.063**
		(0.022)		(0.023)		(0.026)
Other situations		0.063***		0.014		0.011
		(0.022)		(0.024)		(0.026)
Preschool between 3 and 5		−0.072***		−0.034***		−0.054***
		(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. 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, and preschool before 3 years old. Standard errors clustered at school level in parentheses.

*Significant at 10%. **Significant at 5%. ***Significant at 1%.

made to use English as the medium of instruction comes at the expense of a worsening in the results of standard examinations of that subject in Spanish.[15]

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 nontreated periods and schools. Table 10 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.[16] 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, because 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 2004/2005 in posttreatment period, that is, the effect of the program we want to estimate. As with estimates with covariates in Table 8, we only find significant effects for General Knowledge. However, these estimates by groups have the following features: 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 nonsignificant, 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.

15. In a sense, there is a confound, because it is possible that the students do not know less but simply they do not know how to express it in Spanish. But, even if that is the case, this would also suggest that the level of linguistic competence in English is not enough to leap through that barrier. And, possibly more importantly, other standardized examinations which, unlike the CDI, do have academic consequences (at the end of secondary and at the entry to university) are in Spanish, so a negative result in CDI is still an outcome of interest.

16. 11–12 years is the theoretical age that corresponds to sixth grade, which is the grade at which the CDI exam is taken (see Section III.A).

B. Further Look to the Potential Selection Problem

Estimates of δ in Tables 8 and/or 10 will capture the effect of the program not only if there is only selection on observables but also if the selection on unobservable characteristics is highly correlated with the observables that we have, where by selection on unobservables in our model we mean that the diff-in-diff changes in unobservables are endogenous. In the latter case, the β coefficients of the x variables (like parent's education) will be capturing the effect of the unobservables (like educational resources at home) leaving the estimate of the effect of the program ($\widehat{\delta}$) approximately unbiased.[17] 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 nontreated students. Even though the beginning of the program was not anticipated, the treatment lasted for 6 years until we observed our outcome variable, and during that period the following movements of students may occur because of the program:

1. In the Spanish education system, the students who perform badly can be retained in a grade once during primary education. This happens on average to around 15% of the students in any cohort of sixth-grade students.[18] As a consequence of the learning challenges added by the bilingual program, there could be more students retained in a grade than in the previous cohorts in the same school. We would not observe the outcome for these retained students in their cohort because they are not yet in the sixth grade, making the estimated effect more positive than what actually is. If this were the case, in the second-treated graduating class we would observe a higher proportion of retained students than in the last nontreated group at bilingual schools. In our data, the proportion of retained students in the second-treated cohort (2005/2006) is 18.00%, and that proportion in the last nontreated cohort (2003/2004) is 16.53 %. The difference in these

17. Even if one expects the treatment indicator to be correlated with unobservables, the fact that student's school choice was made prior to the announcement of the program, and that changing school is difficult afterwards, will likely make the resulting correlation between ($year2010 \cdot bil$) and the changes in unobservables much smaller than the correlation between x and the unobservables.

18. In our dataset, we define retention as being in an older group than the age that sixth grades should have according to the compulsory schooling rules if not retained.

two proportions is very small and not statistically different from zero—the *p* value is .35 —even if we test it after controlling for observable changes in the composition of the two cohorts. Therefore, this does not seem to be a problem.

2. If a student starting primary education in 2003/2004 was retained in a grade in a bilingual school, he would have gone from a nontreated cohort to a bilingual one. Most of the classmates of that child would have started school in 2004/2005 and, therefore, they would have already participated in the bilingual program for some years. These retained students may have preferred, or may have been recommended to move to a school without the bilingual program in the grade they had to repeat. If this is the case, the treated cohort for which we observe our outcome variable may have a smaller proportion of these retained students from earlier, untreated, cohorts. Looking at our data, we find that for the first group of bilingual schools the proportion of retained students taking the CDI exam falls from 16.53% in 2009 (the last nontreated cohort) to 11.98% in 2010 (the first-treated cohort). The difference is significantly different from zero at 1%. One would expect this factor to improve the outcomes of the treated schools. However, this problem can be solved by comparing the results in the diff-in-diff only for the nonretained students in both the control and treated groups, as we do in Table 9.[19]

3. Some students that were in a bilingual school when the program was implemented might have disliked the program and they could have decided to change school at any point between the year of introduction of the program and the outcome we observe. We conjecture that there will be a very small proportion of students in this group. The reason for our conjecture is that if they had decided to move, they could not have gone to a highly demanded school, since at this stage those schools would have all their vacancies filled. Nevertheless, we do not have individual data to support our guess.

4. Finally, other endogenous movements can be related to the fact that some of the treated schools had vacancies. As mentioned in Section III.B, 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. Because the treatment we evaluate started 6 years before we measure the outcome, new students could have been coming for these reasons for 5 years.[20] Once controlled for retention as indicated, this seems to be a major source of the changes in student population in the bilingual schools reported in Table 2.

TABLE 9

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	−0.117	−0.107	36.36%
	(0.096)	(0.128)	(0.134)	
Postcompulsory secondary	−0.083	−0.210	−0.259**	19.11%
	(0.121)	(0.136)	(0.120)	
Compulsory education or less	−0.115	−0.062	−0.338**	12.33%
	(0.081)	(0.134)	(0.154)	

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., nonimmigrants), not older than 12 years and that do not have special education needs. They are divided by parents' education into three groups. Proportion is the percentage 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%.

Another way to control for the endogenous incoming students to bilingual schools is to use information on who was at those schools before the program was announced. That information is equivalent to an assignment to treatment indicator in experimental programs. For those students in bilingual schools taking the exam in 2009/2010 (i.e., the treated cohort), we know who was already at this school when they were 5 years old. For these students, the implementation of the program was not known when deciding to enroll in this school. We use this information to perform the following two estimates.

19. Another reason to exclude retained students from the comparison is that even if they take the CDI exam with a treated cohort, they have not received full treatment because they entered the program only after being retained; and retention usually does not take place in the first years of primary education.

20. 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 (e.g., due to migration) and we control for this by observing the same school before the program.

TABLE 10

Diff-in-Diff with and without Covariates: Bilingual Schools with More Than 16% of the Students Coming to the School After Age 5 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)
Posttreatment period	−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 2004/2005	−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 2004/2005 in posttreatment period	0.077	−0.017	0.028	−0.092	−0.155	−0.273*
	(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. Although not reported, estimates with *x* include the same covariates as in Table 8.

*Significant at 10%. **Significant at 5%. ***Significant at 1%.

First, we can use that information to detect bilingual schools with a very large proportion of students in the treated cohort who were at the school since they were 5 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 eight bilingual schools that have a proportion of students that were not in that school at 5 years old smaller or equal than 16%. Table 10 presents estimates of Equations (2) and (3) (i.e., diff-in-diff estimates) using as treated group only those eight schools and excluding from the sample the other 17 bilingual schools. The results are similar to the results in Table 8 using the whole sample. The only difference is that the estimated effects are more imprecise as the higher standard errors indicate. 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.

Second, 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 II. 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 schools, we do not have to use the diff-in-diff strategy and we can run a regression using only the 2009/2010 results of the exam. This controls for the selection of schools into the program. To control for selection of students, we include as covariates the characteristics of the students we observe, and we use as an instrumental variable (IV) the indicator of having been at the same school when the student was 5 years old (i.e., having being assigned to treatment).[21] Table 11 contains these two estimates. Both ordinary least squares (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.

In addition to the IV estimation, we have also used the diff-in-diff estimator in the subset of schools that applied for the bilingual education program mentioned in the previous paragraph. The results are in Table 12. Again, the qualitative conclusions are the same as with the entire set of schools.

Falsification test. Finally, in our checks we perform a falsification test with the 2009 and 2010 data, using as (false) treated group the second group of schools implementing the program. Those schools will have their first class of treated students taking the exam in 2011, but in 2010 their students in sixth grade are not yet treated. The schools actually treated in 2010 are excluded for this test. Since the sixth graders in both schools will not be in bilingual programs, there should not be any treatment effect. If we

21. Krueger (1999) is an example in Economics of Education where a variable related to the assignment to treatment is used as instrument to control for potentially endogenous student movements.

TABLE 11
OLS and IV with Schools That Applied to Become a Bilingual School 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 2004/2005 in posttreatment period	−0.070	−0.123	−0.081	−0.046	−0.186*	−0.261**
	(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 educational needs	−0.876***	−0.875***	−0.783***	−0.784***	−0.718***	−0.717***
	(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—university	0.242***	0.243***	0.279***	0.278***	0.232**	0.233**
	(0.086)	(0.085)	(0.101)	(0.100)	(0.093)	(0.093)
Parent education—higher secondary	0.080	0.081	0.210**	0.209**	0.143	0.145
	(0.075)	(0.075)	(0.099)	(0.098)	(0.093)	(0.094)
Parent education—vocational training	0.055	0.057	0.243**	0.241**	0.142	0.145
	(0.102)	(0.102)	(0.116)	(0.114)	(0.107)	(0.107)
Parent education—lower secondary	−0.096	−0.095	0.128	0.127	−0.010	−0.007
	(0.086)	(0.086)	(0.094)	(0.093)	(0.100)	(0.100)
Parent occupation—business, minister, city hall	0.189***	0.190***	0.063	0.062	0.117**	0.120**
	(0.049)	(0.048)	(0.052)	(0.051)	(0.052)	(0.051)
Parent occupation—professional	0.268***	0.268***	0.133***	0.133***	0.088*	0.088**
	(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 at 7 or more	−0.405***	−0.410***	−0.003	−0.000	0.012	0.006
	(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 2009/2010 CDI exam in each of the three subjects. Reference categories for dummies and explanatory variables included in the estimates are as in equations with covariates in Table 8. However, explanatory variables with no significant coefficient in any equation or those variables related to composition of the family living with the student are not reported here. Standard errors clustered at school level in parentheses.

*Significant at 10%. **Significant at 5%. ***Significant at 1%.

find an effect, it could mean that the introduction of the bilingual programs has spillovers to untreated cohorts. More problematic for our estimates, it could also mean that there are pretrends, or that there is some sort of selection of program schools by unobservables. The results of this falsification test are reported in Table 13. The estimated effects for the three subjects are positive but not significantly different from zero, not finding any evidence supporting the aforementioned problems.

C. Results for the Second Cohort of Treated Students in the Schools Selected in 2004/2005

The estimates from Sections IV.A and IV.B 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 2010, this cohort finished sixth grade, the last year of primary education, and took the CDI exam. Likewise, we can use the results of the sixth graders in the CDI exam in 2011 as the output for the second cohort of students treated at those 25 schools. 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.

Table 14 reports the 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 larger (including a less negative effect on General Knowledge) than those

TABLE 12

Diff-in-Diff Estimates Using as Control Group Schools That Applied to Become a Bilingual School and Scored High in the Selection Criteria

	Mathematics	Reading	General Knowledge
Constant	3.759***	3.198***	2.742***
	(0.642)	(0.661)	(0.628)
Year 2010	−0.044	0.013	−0.044
	(0.084)	(0.069)	(0.100)
Bilingual school 2004/2005	0.052	0.201*	0.104
	(0.087)	(0.110)	(0.127)
Bilingual school 2004/2005 in CDI exam 2009/2010	−0.097	−0.203	−0.256*
	(0.109)	(0.123)	(0.152)

Notes: Dependent variables are the individual standardized grades in each of the three subjects. Standard errors clustered at school level in parentheses. Although not reported, estimates include the same covariates as in Table 8.

*Significant at 10%. **Significant at 5%. ***Significant at 1%.

TABLE 13

Falsification Test: Diff-in-Diff Using as False-Treated Group the Schools That Will Implement the Program 1 Year Later

	Mathematics	Reading	General Knowledge
Constant	4.535***	3.116***	3.414***
	(0.133)	(0.123)	(0.138)
Posttreatment period	−0.074***	−0.088***	−0.075***
	(0.012)	(0.013)	(0.015)
Bilingual school 2004/2005	−0.010	−0.076	−0.010
	(0.049)	(0.049)	(0.075)
Bilingual school 2004/2005 in posttreatment period	0.018	0.078	0.075
	(0.057)	(0.051)	(0.079)
Observations	90,178	90,345	90,345

Notes: Dependent variables are the individual standardized grades in each of the three subjects in 2009 and 2010 exams. Standard errors clustered at school level in parentheses. Although not reported, all estimates include the same x covariates as in Table 8.

*Significant at 10%. **Significant at 5%. ***Significant at 1%.

reported in Table 8, 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.

D. Results for the First Cohort of Treated Students in the Schools Selected in 2005/2006

Next, we look at the estimated effects of the first-treated cohort of the 53 schools that became bilingual in 2005/2006. 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. This allows us to check if our results for the schools selected in 2004 to participate are confirmed for the schools selected in 2005, as explained in Section II, there were some significant changes in the selection criteria from one year to the next.

The estimates are reported in Table 15. 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 nonsignificant. This change in the average estimated effect could be because of a composition effect, as the effect is heterogeneous. As seen in Table 9, the effect is higher in absolute value the smaller the level of education of the parents. The students at these 53 schools have better sociodemographic 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 16 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 nonsignificant 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 postcompulsory secondary education the effect of the program in General Knowledge is now not significantly different from zero. The estimated effect is now −0.028 and in Table 9 it was −0.259.[22] Also, all the other estimates for the effect in General Knowledge (column 3 in Table 16) and most of the other estimates in this table are much smaller (in absolute value) than the estimated effects for the first 25 schools.

22. A test of equality of these two estimated effects for the group "postcompulsory secondary" rejects the null hypothesis of equality of effects at 10% (p value is .0588). However, the null hypothesis that the effects for the other groups estimated in Table 16 are equal to those in Table 9 cannot be rejected at typical significant levels (p value is .3001 for "university" and .2320 for "compulsory education or less").

TABLE 14

Diff-in-Diff with and without Covariates: Second Class of Students Treated at the 25 Schools Selected to Implement the Bilingual Program in 2004/2005. Comparing CDI 2010/2011 with CDI 2008/2009

	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)
Posttreatment period	−0.002	−0.034***	−0.004	−0.022*	0.001	−0.016
	(0.012)	(0.012)	(0.014)	(0.013)	(0.014)	(0.014)
Bilingual school 2004/2005	−0.114	−0.017	−0.049	0.041	−0.049	0.075
	(0.074)	(0.059)	(0.097)	(0.092)	(0.093)	(0.094)
Bilingual school 2004/2005 in posttreatment period	0.079	−0.011	0.022	−0.082	−0.076	−0.210***
	(0.078)	(0.086)	(0.097)	(0.096)	(0.090)	(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 2008/2009 and 2010/2011. Standard errors clustered at school level in parentheses. Although not reported, estimates with *x* include the same covariates as in Table 8.

*Significant at 10%. **Significant at 5%. ***Significant at 1%.

TABLE 15

Diff-in-Diff with and without Covariates: First Class of Students Treated at the 54 Schools Selected to Implement the Bilingual Program in 2005/2006

	Mathematics		Reading		General Knowledge	
	No *x*	**With *x***	**No *x***	**With *x***	**No *x***	**With *x***
Constant	0.005	5.177***	0.002	3.265***	0.004	3.720***
	(0.015)	(0.139)	(0.012)	(0.136)	(0.014)	(0.138)
Posttreatment period	−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 2005/2006	−0.088	0.003	−0.042	0.016	−0.024	0.073
	(0.066)	(0.050)	(0.056)	(0.040)	(0.075)	(0.064)
Bilingual school 2005/2006 in posttreatment period	−0.010	−0.058	−0.004	−0.022	−0.029	−0.086
	(0.064)	(0.059)	(0.049)	(0.048)	(0.069)	(0.067)
Observations	109,885	95,861	109,996	96,004	109,996	96,004

Notes: Dependent variables are the individual standardized grades in each of the three subjects in 2009/2010 and 2010/2011. Standard errors clustered at school level in parentheses. Although not reported, estimates with *x* include the same covariates as in Table 8.

*Significant at 10%. **Significant at 5%. ***Significant at 1%.

What can explain the different effects of the program found between the 25 schools selected to implement the program in 2004/2005 and the 53 schools selected in 2005/2006? 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 53 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.[23]

23. We have also done regressions interacting the treatment with the proportion of parents in each of the educational categories and the treatment effect does not change either.

Another explanation could be that those selected in 2005/2006 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 II, in 2005/2006, 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 2005/2006 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 to 2005 is because of 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.

TABLE 16

Diff-in-Diff for the 2005/2006 Schools: Estimated Treatment Effects Using Separate Regressions by Observable Groups of Students

Groups by Parents' Education	Mathematics	Reading	General Knowledge	Proportion
University	−0.099	−0.070	−0.026	37.53%
	(0.076)	(0.062)	(0.077)	
Postcompulsory secondary	−0.006	0.004	−0.028	19.92%
	(0.075)	(0.086)	(0.086)	
Compulsory education or less	−0.057	−0.091	−0.199*	11.76%
	(0.129)	(0.068)	(0.111)	

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., nonimmigrants), 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 percentage 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%.

V. CONCLUDING REMARKS

All our estimates to identify the effect of the bilingual program on different learning outcomes, which control for observable students' characteristics and use several ways to control for self-selection, lead to the same conclusion: there is a clear negative effect, which is quantitatively substantial, on learning the subject taught in English, and the effect is not significantly different from zero on mathematical skills and on reading in Spanish. 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 parents 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 are a stark reminder about the necessity of introducing policies in a way that facilitates 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 especially 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

Admiraal, W., G. Westhoff, and K. de Bot. "Evaluation of Bilingual Secondary Education in The Netherlands: Students' Language Proficiency in English." *Educational Research and Evaluation*, 12, 2006, 75–93.

Anghel, B., and A. Cabrales. "The Determinants of Success in Primary Education in Spain." *Journal of Public Programs and Policy Evaluation*, 2, 2014, 22–53.

Angrist, J. D., and V. Lavy. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114, 1999, 533–75.

Banfi, C., and R. Day. "The Evolution of Bilingual Schools in Argentina." *International Journal of Bilingual Education and Bilingualism*, 7, 2004, 398–411.

Bentolila, S., and A. Ichino. "Unemployment and Consumption Near and Far Away from the Mediterranean." *Journal of Population Economics*, 21, 2008, 255–80.

Bergin, A., T. Conefrey, J. FitzGerald, and I. Kearney ."Recovery Scenarios for Ireland: An Update." Mimeo, 2009.

Center for Applied Linguistics. "Directory of Foreign Language Immersion Programs in U.S. Schools." 2011. Accessed December 22, 2013. http://www.cal.org/resources/immersion/.

Fidrmuc, J. and J. Fidrmuc. "Foreign Languages and Trade: What Are You Sinking About?" Mimeo, 2009.

Fry, R., and L. B. Lowell. "The Value of Bilingualism in the U.S. Labor Market." *Industrial & Labor Relations Review*, 57, 2003, 128–40.

Ginsburgh, V. A., and J. Prieto-Rodríguez. "Returns to Foreign Languages of Native Workers in the EU." *Industrial & Labor Relations Review*, 64, 2011, 599–618.

Greene, J. A. "A Meta-Analysis of the Effectiveness of Bilingual Education." University of Texas at Austin, Mimeo, 1998.

Hofstetter, C. H. "Contextual and Mathematics Accommodation Test Effects for English-Language Learners." *Applied Measurement in Education*, 16(2), 2003, 159–88.

Krueger, A. B. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, 114, 1999, 497–532.

Jepsen, C. "Bilingual Education and English Proficiency." University of Kentucky, Mimeo, 2009.

Munshi, K., and M. Rosenzweig. "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy." *American Economic Review*, 96(4), 2006, 1225–52.

Ordóñez, C. L. "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, 2004, 449–74.

Saiz, A., and E. Zoido. "Listening to What the World Says: Bilingualism and Earnings in the United States." *Review of Economics and Statistics*, 87, 2005, 523–38.

Williams, D. R. "Multiple Language Usage and Earnings in Western Europe." *International Journal of Manpower*, 32, 2011, 372–93.

Willig, A. C. "A Meta-Analysis of Selected Studies on the Effectiveness of Bilingual Education." *Review of Educational Research*, 55, 1985, 269–317.