UNIVERSIDAD CARLOS III DE MADRID

working papers

Working Paper 12-34
Economic Series
December, 2012

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

Publicizing the results of standardized external tests: Does it have an effect on school outcomes?*

Brindusa Anghel
FEDEA
Antonio Cabrales
Universidad Carlos III de Madrid
Jorge Sainz
Universidad Rey Juan Carlos
Ismael Sanz
Universidad Rey Juan Carlos
December 2012

Abstract

We study the effect of standardized external tests on students' academic outcomes. We exploit the fact that only one of the 17 Spanish regions started doing and publishing the results of standardized tests in 2005 to apply a difference-in-difference methodology, using outcomes of the PISA study from 2000 to 2009. We later confirm our results using synthetic control methods. Using data from a single country allows us to minimize biases arising from differences in legal frameworks, social or cultural environments. Our econometric analysis lends plausibility to the hypothesis that this type of test significantly improves student outcomes. A key novelty is that our exams do not have academic consequences for the students, so that effects have to come directly from the impact on teachers and administrators.

Keywords: external and standardized tests, PISA, difference-in-difference, synthetic control methods

* We gratefully acknowledge the help of Javier Garcia Crespo and Guillermo Gil to obtain the data and the support from the Ministry of Science and Technology from grants ECO2012-31985 (Anghel).

1. Introduction

External standardized tests allow the administration to better monitor the education process and outcome of the schools. In most of the countries that have these tests, the results of the exam are public and can be used by parents to make decisions.[1] This closer monitoring by parents and administrators provides an additional motivation for teachers and principals to improve the education results of their students. This potential for improvement has encouraged an increasing number of countries to use external examinations as a tool to increase accountability. The Programme for International Student Assessment (PISA) report (OECD, 2010) documents the fact that 22 out of the 34 OECD countries have introduced standards-based external examinations in a majority of their schools. Two more countries, Germany and the US, have this type of tests only in some of their Ländern and states. All in all, two thirds of the 15 years old OECD students attend schools in which there is an external and standardized test.

The existing empirical evidence is supportive of the hypothesis that countries with external exit-exam systems have a better performance in international student achievement tests. The first evidence for this was given by Bishop (1997) for students doing the 1991 IAEP math, science, and geography tests and Bishop (2006) with the PISA 2000 results. Overall, the existing cross-country evidence suggests that the effect of external exit exams on student achievement may well be half or more of a grade-level equivalent, or between 20 and 40 percent of a standard deviation of the respective international tests (OECD, 2010 and 2012 and Hanushek and Woessmann, 2011).

This evidence has been criticized on two grounds. First, these studies use cross-sectional data and therefore the adoption of testing by a country is endogenous, and unobserved heterogeneity could bias the results. Second, the introduction of external tests may lead to "teaching to the test". However, some studies have found the same positive association between central exams and student achievement within countries where some regions have external exam systems and others do not have them.[2] This evidence rules out the possibility that unobserved national-level factors correlated with the existence of tests drive the observed positive correlation between those tests and students' outcomes. In addition, students in countries with national external exams have been found to achieve

[1] Even in countries where school zones comprise a single school, concerned parents can decide where to live using school quality as an input to their choice.
[2] See, e.g. Bishop (1997) for Canadian provinces, Jürges et al. (2005) and Woessman (2010) for Germany, and Bishop et al. (2001), for US states.

better results in other international tests such as PISA, PIRLS or TIMMS. To the extent that those tests are different in nature from national ones, this may rule out that "teaching to the test" is a main factor driving the better outcomes of students in countries or regions with national external exams.

A different difficulty of earlier studies is that they are not very clear on what are the channels through which exit exams are effective. This is because for the most part these exams have academic consequences for the students, thereby providing reasons for improvement both to the professionals and to the students. The present study uses a special feature of the Spanish education system to tease out school and student incentives, while at the same time controlling for biases arising from unobserved national-level heterogeneity and arguably also "teaching to the test".

The special feature to which we refer earlier is that the main Spanish education law (Ley Orgánica de la Educación, LOE 2006) allows the Regions to conduct education system assessments *as long as the results are not used for grading students or ranking schools* (article 140). That means Spanish exams are not "Curriculum-Based External Exit Examination (CBEEE)" as defined by Bishop (1997), because such examinations should "offer signals of student accomplishments that have real consequences for the student and define achievement relative to an external standard, not relative to other students in the classroom or the school". This means that the effects of such exams in Spain, if any, have to come directly only from changes in incentives for schools, although in the end those can, and probably will, have an impact on the students' efforts.

The Region of Madrid introduced a standard external test, called *"prueba de Conocimientos y Destrezas Indispensables"* (also known for short as *CDI* test) which means "Indispensable Knowledge and Skills exam" in the academic year 2004/05. The grade achieved by the student in this exam does not have "real academic consequences" for most students, so it cannot be considered a CBEEE.[3] So the effects of this initiative will necessarily go directly only through changes in teacher motivation. The Region of Madrid is also the only one that publishes and makes available the average results of each of the schools in the external test to the public opinion.[4]

[3] A student with good grades in compulsory secondary schooling and a good mark in the CDI test obtains a certification with Merit or with Distinction, rather just a Certificate, but this has no implications for admissions to schools beyond the compulsory schooling, or for grants, nor is there evidence that employers look at those distinctions. For students with really extraordinary grades (only 25 a year in a region with over 50,000 students in the last year of compulsory secondary schooling) they can obtain an Extraordinary Award yielding a cash prize of 1,000 Euros and a trip to a "cultural destination".

[4] Other regions have external standardized exams where all schools are tested, but Madrid is the only one publishing the results.

All the regions in Spain operate under the same legal framework regulating the principles, objectives, and organization of the different school levels (pre-primary, primary, compulsory secondary, post-compulsory secondary), as well as up to 65% (55% in historical regions) of the contents and subjects studied. Hence, along with the amount of public financing of schools, for which we can control, the other main observable difference in education between Spanish regions is the appearance in the period of study of this standardized external exam in Madrid whose results are published.

This feature allows us to conduct a difference-in-difference (diff-in-diff) analysis comparing the PISA results of the treated region (Madrid) before and after the CDI test was introduced with the rest of Spanish regions before and after the treatment. This diff-in-diff approach allows us to control for the unobservable time-invariant factors affecting Madrid. Dealing with regions of the same country we also exclude some unobservable effects that appear in cross-country studies with different legislations and cultures.

The fact that we are analyzing a single country also allows us to apply the new inferential methods of synthetic control for comparative case studies proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010). We use a combination of other Spanish regions to construct a *synthetic control region*, which resembles relevant education characteristics of Madrid before the introduction of the CDI test. The subsequent education outcome evolution of this "counterfactual" Madrid without CDI is compared to the actual experience of Madrid. The idea behind the synthetic control approach is that a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone. Transparency and safeguard against extrapolation are two attractive features of the synthetic control method relative to traditional regression methods.

Our results are also more protected than others from the critique that they are achieved by "teaching to the test". This is because our measure of outcome, namely, the results in the PISA exam, have somewhat distinct objectives and measure different things than the CDI exam in whose effect we are interested. The Madrid CDI exam questions evaluate knowledge and they are directly related to material seen in Language and Mathematics classes during the academic year. In contrast, the PISA exam questions (called stimulus) are more related to cognitive processes (access and retrieve; integrate and interpret; reflect and evaluate) and on how to use knowledge in particular contexts. That is the PISA evaluation is more related to competencies whereas the Madrid CDI is more related to knowledge.

The paper is organized as follows. Section 2 describes in some detail the institutional setup and the CDI external and standard test. Section 3 discusses the data. Section 4 discusses the econometric methodology and it contains the main results of the paper. Section 5 shows the results of the synthetic control methods. Section 6 concludes.

2. Institutional Setup

The Madrid regional government has been conducting since the academic year 2004/05 a standardized external exam for all 6th grade students in the region, who are hence in the final year of primary school (around 11-12 years old). Three years later, the region introduced another standardized and external exam in the 9th grade (the third year of the secondary school, which is the last common academic year for the students). These exams are compulsory for all primary and secondary schools (public or private). The exam measures what the authorities consider basic knowledge in mathematics (exercises and problems) and language (dictation, reading, general knowledge and questions related to a text).

Our aim is to test whether the introduction of these exams has improved the academic outcomes of the students in Madrid. We use as a measure of student achievement the scores of the exams conducted for the OECD Programme for International Student Assessment (PISA). PISA analyses the key competencies in reading, mathematics and science of 15-year-old students in OECD member countries and partner countries/economies through its triennial surveys. The metric for the overall reading scale is based on a mean for OECD countries set at 500 in PISA 2000, with a standard deviation of 100. PISA conducted its first tests in 2000 covering reading as a major assessment area, and providing a summary profile of the skills of mathematics and science. In 2003, mathematics was the main focus and in 2006 it was science. In 2009, PISA started another cycle, focusing on reading again. When an area is the main focus of the exam two thirds of the exam time is devoted to this area, allowing for its deeper analysis. Since both PISA 2000 and PISA 2009 focused on reading, it is possible to obtain very detailed comparisons of how student performance in that area changed over that period. Comparisons over time in the areas of mathematics and science are somewhat more limited.

In the PISA test, each participating student spends two hours carrying out pencil-and-paper tasks in reading, mathematics and science. The assessment includes tasks requiring students to construct their own answers as well as multiple-choice questions. In addition, students also answer a survey that takes about 30 minutes to complete and that includes questions about their personal background.

3. Description of the Data

The first CDI exam took place in the academic year 2004/05, so we consider this as the year in which the treatment (the introduction of a standardized exam) was first implemented. For this reason, we compare the results of students of the region of Madrid in (i) Reading, using PISA 2000 and PISA 2009 and (ii) in Mathematics, using PISA 2003 and PISA 2009.[5]

Our first methodology for analysis will be a diff-in-diff regression approach. We construct the treatment and the control groups in the following way: the treatment group before the treatment (the introduction of the CDI exam) is the group of students from the region of Madrid who took the PISA exam in 2000 for Reading or 2003 for Mathematics, the treatment group after the treatment is the group of students who took the PISA exam in 2009, and the control group is formed by students from the other regions of Spain before (PISA 2000 or 2003) and after the treatment (PISA 2009).

The PISA questionnaire allows us to control for various student, family and school characteristics. The student and family characteristics are: gender, age, nationality (immigrant or Spanish), parents' nationality, languages other than Spanish spoken at home, structure of the family (single parent family, nuclear family, mixed family), learning time in hours per week in Reading and Mathematics (hours per week in Language or Mathematics courses), the index of economic, social and cultural status (ESCS index) calculated by OECD.[6] The school characteristics are the type of school (public, charter or private), the location of the school (village, small town, town, city or large city), student/teacher ratio, school size, whether the school uses assessments to compare to district/national performance, whether the school uses assessments to make judgments about teacher's effectiveness, the proportion of girls in the school, the school average of ESCS index, the percentage of immigrant students in school, and school average learning time in Reading and Mathematics.

5 We will not use the PISA scores in Science, since the first year Science was the main focus was 2006 and this is after our treatment was applied.

6 The PISA index of economic, social and cultural status (ESCS) was derived from the following three indices: highest occupational status of parents, highest educational level of parents in years of education according to ISCED and home possessions (OECD, 2010).

The tables below contain the descriptive statistics of these four groups, for the most relevant characteristics of students and schools. Table 1 describes the treatment and the control groups in PISA 2000 and in PISA 2009 in Reading and Table 2 describes the two groups in PISA 2003 and in PISA 2009 in Mathematics.

Table 1 – Descriptives statistics of students and schools in PISA 2000 and PISA 2009 (Reading)

	Treatment group before change - School Madrid in PISA exam 2000				Control group before change - School NO Madrid in PISA exam 2000				Treatment group after change - School Madrid in PISA exam 2009				Control group after change - School NO Madrid in PISA exam 2009			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Subjects - Plausible Values																
Reading - PV1	510.16	79.75	245.12	697.46	491.56	84.70	150.64	777.08	504.38	85.14	74.61	721.54	483.75	89.49	6.65	814.71
Reading - PV2	509.94	82.20	204.02	711.64	491.47	84.37	151.55	756.66	504.39	85.52	91.12	822.19	483.82	89.90	60.42	824.33
Reading - PV3	510.29	81.44	237.66	716.00	491.89	84.88	157.91	768.72	503.43	83.68	60.18	733.45	483.84	89.65	19.36	814.71
Reading - PV4	508.55	81.72	229.34	710.26	491.36	84.87	117.01	751.45	503.58	85.14	80.22	702.07	483.67	90.16	29.15	877.24
Reading - PV5	508.86	81.96	176.62	709.59	491.23	85.21	112.65	782.35	504.67	84.54	124.31	729.32	483.96	89.75	64.42	904.50
Individual characteristics																
Female	0.51	0.50	0	1	0.51	0.50	0	1	0.50	0.50	0	1	0.49	0.50	0	1
Age	15.80	0.28	15.33	16.25	15.79	0.28	15.33	16.25	15.87	0.28	15.33	16.33	15.86	0.29	15.33	16.33
Immigrant	0.03	0.18	0	1	0.02	0.15	0	1	0.16	0.37	0	1	0.09	0.28	0	1
Mother immigrant	0.07	0.25	0	1	0.04	0.19	0	1	0.21	0.41	0	1	0.12	0.33	0	1
Father immigrant	0.05	0.22	0	1	0.03	0.18	0	1	0.20	0.40	0	1	0.11	0.31	0	1
Languages other than Spanish spoken at home	0.01	0.09	0	1	0.17	0.37	0	1	0.05	0.22	0	1	0.16	0.37	0	1
Single parent family	0.17	0.37	0	1	0.17	0.37	0	1	0.16	0.37	0	1	0.13	0.34	0	1
Nuclear family	0.79	0.41	0	1	0.79	0.41	0	1	0.83	0.38	0	1	0.85	0.35	0	1
Mixed family	0.04	0.21	0	1	0.05	0.21	0	1	0.01	0.09	0	1	0.01	0.11	0	1
Learning time (hours/week) in Language	3.30	0.74	0.00	4.58	3.04	0.74	0.00	4.33	3.66	0.68	1.67	9.00	3.37	0.70	0.00	9.00
ESCS (Index of Economic, Social and Cultural Status)	-0.16	1.01	-2.79	2.13	-0.40	1.05	-4.05	2.21	-0.09	1.05	-3.40	2.85	-0.26	1.05	-5.34	3.41
ESCS squared	1.05	1.14	0.00	7.80	1.26	1.42	0.00	16.42	1.11	1.30	0.00	11.57	1.18	1.44	0.00	28.54
School characteristics																
Public school	0.56	0.50	0	1	0.63	0.48	0	1	0.60	0.49	0	1	0.63	0.48	0	1
Private school	0.16	0.36	0	1	0.08	0.27	0	1	0.08	0.27	0	1	0.03	0.18	0	1

Charter school	0.28	0.45	0	1	0.29	0.46	0	1	0.32	0.47	0	1	0.34	0.47	0	1
School in village	0.00	0.00	0	0	0.02	0.15	0	1	0.02	0.14	0	1	0.06	0.24	0	1
School in small town	0.00	0.00	0	0	0.24	0.42	0	1	0.11	0.32	0	1	0.26	0.44	0	1
School in town	0.18	0.38	0	1	0.34	0.47	0	1	0.27	0.44	0	1	0.35	0.48	0	1
School in city	0.36	0.48	0	1	0.36	0.48	0	1	0.16	0.36	0	1	0.32	0.47	0	1
School in large city	0.47	0.50	0	1	0.04	0.21	0	1	0.44	0.50	0	1	0.01	0.12	0	1
Student/teacher ratio	16.93	4.39	10.90	26.80	14.15	4.66	5.66	27.60	12.76	4.23	1.18	20.27	11.29	4.69	0.82	39.88
School size	1,048.43	493.49	335.00	2,139.00	725.64	345.15	93.00	1,742.00	851.82	416.07	100.00	2,268.00	685.72	392.01	44.00	2,785.00
Assessments used to compare the school to district/national performance	0.32	0.47	0	1	0.18	0.39	0	1	0.50	0.50	0	1	0.28	0.45	0	1
Assessments used to make judgements about teacher's effectiveness	0.47	0.50	0	1	0.40	0.49	0	1	0.62	0.48	0	1	0.45	0.50	0	1
No. of observations – schools	**20**				**165**				**51**				**838**			
No. of observations – students	**679**				**5535**				**1453**				**24434**			

Table 2: Descriptive statistics of students and schools in PISA 2003 and PISA 2009 (Mathematics)

	Treatment group before change - School Madrid in PISA exam 2003				Control group before change - School NO Madrid in PISA exam 2003				Treatment group after change - School Madrid in PISA exam 2009				Control group after change - School NO Madrid in PISA exam 2009			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Subjects - Plausible Values																
Mathematics - PV1	493.98	92.37	238.46	706.68	495.25	85.81	122.01	807.17	497.47	88.34	179.79	773.49	490.28	93.53	48.07	803.25
Mathematics - PV2	492.34	91.10	197.49	701.70	495.11	86.17	56.27	788.63	497.52	88.68	205.49	820.23	490.47	93.75	43.39	821.16
Mathematics - PV3	494.73	87.87	216.34	694.69	494.23	85.59	137.67	793.61	498.24	88.22	222.63	749.35	490.84	93.91	4.45	778.01
Mathematics - PV4	495.29	90.48	229.35	731.61	495.14	85.91	117.34	770.01	496.53	87.95	196.22	740.23	490.26	94.10	3.67	823.97
Mathematics - PV5	491.91	92.66	202.94	723.35	494.47	85.62	143.82	777.49	497.89	87.76	193.03	777.39	490.26	93.54	5.23	818.05
Individual characteristics																
Female	0.50	0.50	0	1	0.51	0.50	0	1	0.50	0.50	0	1	0.49	0.50	0	1
Age	15.85	0.29	15.25	16.33	15.86	0.29	15.25	16.42	15.87	0.28	15.33	16.33	15.86	0.29	15.33	16.33
Immigrant	0.07	0.25	0	1	0.03	0.17	0	1	0.16	0.37	0	1	0.09	0.28	0	1
Mother immigrant	0.12	0.33	0	1	0.04	0.20	0	1	0.21	0.41	0	1	0.12	0.33	0	1
Father immigrant	0.09	0.29	0	1	0.04	0.19	0	1	0.20	0.40	0	1	0.11	0.31	0	1
Language other than Spanish spoken at home	0.04	0.19	0	1	0.14	0.35	0	1	0.05	0.22	0	1	0.16	0.37	0	1
Single parent family	0.18	0.38	0	1	0.13	0.33	0	1	0.16	0.37	0	1	0.13	0.34	0	1
Nuclear family	0.79	0.41	0	1	0.83	0.38	0	1	0.83	0.38	0	1	0.85	0.35	0	1
Mixed family	0.02	0.15	0	1	0.02	0.15	0	1	0.01	0.09	0	1	0.01	0.11	0	1
Learning time (hours/week) in Mathematics	3.05	0.60	0.92	7.33	2.97	0.73	0.00	15.00	3.20	0.81	1.67	9.00	3.44	0.67	0.00	9.00
ESCS (Index of Economic, Social and Cultural Status)	-0.12	0.97	-3.74	2.10	-0.20	0.98	-3.40	2.39	-0.09	1.05	-3.40	2.85	-0.26	1.05	-5.34	3.41
ESCS squared	0.94	1.38	0.00	13.96	1.00	1.28	0.00	11.55	1.11	1.30	0.00	11.57	1.18	1.44	0.00	28.54
School characteristics																
Public school	0.54	0.50	0.00	1.00	0.55	0.50	0.00	1.00	0.60	0.49	0.00	1.00	0.63	0.48	0.00	1.00
Private school	0.07	0.26	0	1	0.06	0.24	0	1	0.08	0.27	0	1	0.03	0.18	0	1

Charter school	0.39	0.49	0	1	0.39	0.49	0	1	0.32	0.47	0	1	0.34	0.47	0	1
School in village	0.00	0.00	0	0	0.05	0.22	0	1	0.02	0.14	0	1	0.06	0.24	0	1
School in small town	0.05	0.23	0	1	0.24	0.43	0	1	0.11	0.32	0	1	0.26	0.44	0	1
School in town	0.23	0.42	0	1	0.33	0.47	0	1	0.27	0.44	0	1	0.35	0.48	0	1
School in city	0.13	0.33	0	1	0.36	0.48	0	1	0.16	0.36	0	1	0.32	0.47	0	1
School in large city	0.59	0.49	0	1	0.02	0.15	0	1	0.44	0.50	0	1	0.01	0.12	0	1
Student/teacher ratio	15.61	5.48	9.44	24.36	13.23	5.31	1.38	44.30	12.76	4.23	1.18	20.27	11.29	4.69	0.82	39.88
School size	880.37	401.45	389	1951	710.70	432.90	95	2819	851.82	416.07	100	2268	685.72	392.01	44	2785
Assessments used to compare the school to district/national performance	0.12	0.33	0	1	0.26	0.44	0	1	0.50	0.50	0	1	0.28	0.45	0.00	1.00
Assessments used to make judgements about teacher's effectiveness	0.24	0.42	0	1	0.40	0.49	0	1	0.62	0.48	0	1	0.45	0.50	0.00	1.00
No. of observations - schools	**18**				**365**				**51**				**838**			
No. of observations - students	**511**				**10280**				**1453**				**24434**			

The two tables show a very similar evolution of the characteristics of students and schools when we compare PISA 2000 and PISA 2009 in Reading and when we compare PISA 2003 and PISA 2009 in Mathematics.

If we compare the treatment group and the control group before and after the change in Reading we can see patterns that are very similar across both groups: the proportion of girls and students coming from single parent family decreases slightly whereas age, immigrants, learning time, and the Index of Economic, Social and Cultural Status (ESCS) increases. This is consistent with the fact that Spain has experienced a large inflow of immigrants in the last decade and had converged towards the EU and OECD GDP per capita, a process that has, since 2009, reversed.[7] Nevertheless, the rise in the share of immigrants between 2000 and 2009 was higher in the region of Madrid (in Reading, from 3% to 16%) than in the rest of regions (in Reading, from 2% to 9%). In addition, the share of students speaking foreign languages other than Spanish increased in Madrid over the period (from 1% to 5%) whereas it remained constant in the control group (17% versus 16%).

If we look at the school characteristics, we observe a decrease in the number of private schools and an increase in the number of charter schools over the period 2000-2009. This could be due to the fact that some private schools have demanded and achieved from the public administration their transformation into charter schools, thus lowering the fees to be paid by the student's families and avoiding losing enrolment. Nevertheless, the official data from the Statistical Office of the Spanish Ministry of Education, Culture and Sports shows that the rise in the students of charter schools has come from a reduction in the number of students in the public schools. This is in contrast of the PISA sample, which shows that the rise of the students in charter schools come from a reduction in the private schools. That is, the PISA coverage of private schools decreased from 2000 to 2009 whereas the coverage of public schools increased. This could be explained by the fact that the sample of schools in cities or large cities in 2009 decreased whereas those in towns and villages increased.

Student/teacher and school size ratio decreased along the two periods. We also observe that the percentage of schools, which declare that they carry out assessments used to compare the school to district/national performance or assessments used to make judgements about teacher's effectiveness increased during the two periods and in both the

[7] Spain went from less than 1% of immigrants in the population to almost 10% during this period. The Spanish GDP per capita in PPS terms increased from 97% of that of the EU-27 in 2000 to 103% in 2009 (Source: Eurostat).

control and the treatment group. In summary, the descriptive statistics show that the trends that we observe when we compare the treatment and control group are similar.

The control group and the treatment group have also similar patterns in the PISA 2003 and PISA 2009, the years we are using for the Mathematics analysis. The only exceptions are the proportion of girls and, above all, the ESCS index. In both cases, the indicator of the Region of Madrid increased whereas the one of the control group slightly decreased.

4. Econometric Methodology and Results

In order to estimate the impact of the introduction of a standardized exam in the region of Madrid on students' outcomes, we propose a diff-in-diff approach. We use as the outcome for student performance, the PISA scores of students. These are calculated using imputation methods, denoted plausible values (OECD, 2009). Thus, for a given year, the score of student *i* in school *j* is given by:

$$y_{ij} = \alpha_0 + \alpha_1 \mathrm{Madrid_j} + \alpha_2 \mathrm{PISA2009} + \delta \mathrm{Madrid_j} * \mathrm{PISA2009} + \beta_{\mathrm{i}} x_i + \beta_{\mathrm{j}} x_j + v_j + u_i + \varepsilon_{ij}$$

where x_i are observable characteristics of students and their families described above, x_j are observable characteristics of schools, $Madrid_j$ is a dummy variable for the schools located in the region of Madrid (i.e. it takes the value 1 for the treated group), $PISA\mathbf{2009}$ is a dummy variables for students who took the PISA exam in 2009 (after the introduction of the standardized exam in the region of Madrid), $Madrid_j * PISA_{2009}$ indicates whether school j is in the region of Madrid and participated in PISA exam 2009 (i.e. it takes the value 1 for the treated group after the treatment), u_i are unobservable characteristics of students, such as effort or ability, v_j are unobservable characteristics of school, like quality of the Principal and teachers, and ε_{ij} is a random shock.

Our parameter of interest is δ, corresponding to the variable $Madrid_j * PISA_{2009}$, which coincides with the introduction of a standardized exam (the CDI exam) in the region of Madrid.

Tables 3 and 4 below show the results of the diff-in-diff estimation for Reading, using the samples of students from Spain in PISA 2000 and PISA 2009, and for Mathematics, using the samples of students from Spain in PISA 2003 and PISA 2009. Since PISA database provides five plausible values, which are allocated to each student, we use the methodology proposed by the OECD for the computation of regression coefficients and their respective standard errors. According to OECD (2009), statistical analyses should be performed independently on each of these five plausible values and results should be aggregated to obtain the final estimates of the statistics and their respective standard errors.

Table 3: Difference in difference estimations for PISA 2000 and PISA 2009 in Reading

Variable	(1)	(2)	(3)	(4)
Madrid	13.620*	9.327	2.071	2.761
	(8.107)	(5.841)	(8.001)	(6.869)
Madrid*PISA2009	**12.267**	**13.718****	**16.985****	**15.415****
	(9.703)	**(6.867)**	**(8.083)**	**(7.657)**
PISA2009	-13.185***	-9.499***	-5.448	-8.436**
	(3.635)	(2.995)	(3.843)	(3.744)
Student characteristics				
Female		27.230***	26.115***	26.231***
		(1.789)	(1.781)	(1.696)
Age		18.521***	19.680***	19.180***
		(3.118)	(3.372)	(3.366)
Immigrant		-17.608***	-16.988**	-19.294***
		(6.301)	(6.778)	(6.504)
Mother immigrant		-13.890**	-14.851**	-11.932*
		(6.766)	(7.345)	(7.027)
Father immigrant		-8.080	-7.079	-5.043
		(5.609)	(5.983)	(6.043)
Foreign language spoken at home		-4.108	-2.888	-0.192
		(2.689)	(2.615)	(2.721)
Single parent family		-4.899**	-5.730**	-6.146***
		(2.288)	(2.312)	(2.178)
Mixed family		-22.204***	-24.441***	-25.051***
		(5.404)	(5.750)	(5.911)
Learning time (hours/week) in Reading		-8.492***	-8.690***	-12.495***
		(1.420)	(1.545)	(1.587)
ESCS (Index of Economic, Social and Cultural Status)		27.955***	24.845***	19.628***
		(1.100)	(1.072)	(0.852)
ESCS squared		-1.564**	-1.868**	-1.656**
		(0.757)	(0.763)	(0.710)
School characteristics				
Private school			23.473***	-3.719
			(8.548)	(8.095)
Charter school			12.103**	-1.032
			(5.776)	(5.099)
School in small town			1.078	1.581
			(6.443)	(4.890)
School in town			1.749	0.086
			(6.653)	(5.042)

School in city			7.013	4.163
			(7.233)	(5.556)
School in large city			12.746	8.784
			(8.965)	(6.671)
Student/teacher ratio			0.038	0.373
			(0.657)	(0.544)
School size			0.005	-0.006*
			(0.004)	(0.004)
Assessments used to compare the school to district/national performance			-1.527	-2.520
			(2.993)	(2.959)
Assessments used to make judgements about teacher's effectiveness			0.592	-0.442
			(2.669)	(2.389)
Propotion of girls in school				-14.228
				(13.844)
School average of ESCS index				23.880***
				(2.808)
School average of ESCS index squared				-8.226***
				(2.706)
% of Immigrant students in school				0.471
				(0.451)
% of Mother immigrants in school				-0.109
				(0.336)
% of Father immigrants in school				-0.538
				(0.415)
School average learning time (hours/week) in Reading				11.089***
				(3.485)
Constant	490.770***	229.792***	197.296***	221.023***
	(2.793)	(49.645)	(54.414)	(55.028)

Notes:

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

2. Base categories for dummies: male, student Spain, Mother Spain, Father Spain, Nuclear family, Public school, School in village

Table 4: Difference in difference estimations for PISA 2003 and PISA 2009 in Mathematics

Variable	(1)	(2)	(3)	(4)
Madrid	9.580	7.819	-3.420	-6.879
	(9.973)	(6.039)	(7.350)	(6.791)
Madrid*PISA2009	**5.415**	**-0.084**	**4.272**	**7.323**
	(10.559)	**(6.606)**	**(7.268)**	**(7.509)**
PISA2009	-2.296	6.752**	6.619**	8.484**
	(3.360)	(2.688)	(3.227)	(3.514)
Student characteristics				
Female		-13.744***	-14.519***	-15.746***
		(1.626)	(1.605)	(1.584)

Age		15.847***	14.596***	15.139***
		(3.269)	(3.433)	(3.325)
Immigrant		-21.247***	-21.009***	-20.172***
		(6.481)	(6.973)	(6.948)
Mother immigrant		-9.528**	-8.680*	-7.674*
		(4.416)	(4.507)	(4.433)
Father immigrant		-13.285***	-15.604***	-13.873***
		(5.073)	(5.339)	(5.268)
Foreign language spoken at home		2.083	2.103	-0.132
		(2.619)	(2.852)	(2.997)
Single parent family		-8.722***	-9.873***	-10.985***
		(2.771)	(2.723)	(2.699)
Mixed family		-5.591	-10.635	-11.649*
		(5.366)	(6.748)	(6.902)
Learning time (hours/week) in Mathematics		-6.616***	-6.759***	-4.840***
		(1.560)	(1.620)	(1.844)
ESCS (Index of Economic, Social and Cultural Status)		29.471***	26.666***	21.539***
		(1.024)	(1.043)	(0.853)
ESCS squared		-0.177	-0.644	-0.046
		(0.735)	(0.799)	(0.657)
School characteristics				
Private school			14.301**	-0.962
			(6.453)	(6.895)
Charter school			10.068**	2.393
			(4.193)	(4.053)
School in small town			3.890	3.007
			(4.283)	(4.901)
School in town			-1.624	-2.684
			(4.961)	(5.075)
School in city			4.235	1.968
			(5.395)	(5.389)
School in large city			17.099**	15.854*
			(8.373)	(8.311)
Student/teacher ratio			-0.024	-0.161
			(0.441)	(0.446)
School size			0.009*	-0.000
			(0.005)	(0.004)
Assessments used to compare the school to district/national performance			-0.661	-3.048
			(3.322)	(3.235)
Assessments used to make judgements about teacher's effectiveness			-1.618	-1.442
			(3.460)	(3.471)
Propotion of girls in school				49.186***
				(15.062)
School average of ESCS index				22.307***

				(3.094)
School average of ESCS index squared				-10.684***
				(3.716)
% of Immigrant students in school				0.032
				(0.333)
% of Mother immigrants in school				0.106
				(0.342)
% of Father immigrants in school				-0.420
				(0.386)
School average learning time (hours/week) in Mathematics				-1.111
				(4.785)
Constant	483.786***	273.629***	285.233***	281.827***
	(2.370)	(52.352)	(55.280)	(54.504)

Notes:
1. Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%
2. Base categories for dummies: male, student Spain, Mother Spain, Father Spain, Nuclear family, Public school, School in village

The first column of the tables shows the estimation results without any control variables. This would be the raw average effect of our treatment. The second column includes individual characteristics of the students and the third and the forth columns add gradually school characteristics.

When we estimate the diff-in-diff without any covariates, the coefficient for the treatment is not statistically significant for both Reading and Mathematics.

However, results of the diff-in-diff estimation for Reading in Table 3, columns (2)-(4) show a positive and statistically significant effect of our treatment on the PISA scores. In the second column, when we control for individual characteristics of students, the coefficient of the treatment variable is positive and significant.[8] The inclusion of school characteristics in columns (3) and (4) does not change this result. We find a relative improvement in PISA scores in Reading in the region of Madrid between 2000 and 2009 that cannot be explained by observable variables of a magnitude of 14 to 17 PISA points. In 2009, Spain was significantly below the OECD average in Reading by 12 points. If the results are totally explained by the introduction and publication of external exams, this could imply that generalizing those exams would raise the level of Spain in Reading above the OECD average.

In Table 4 we run the same estimations, but for Mathematics and using the scores in PISA 2003 and PISA 2009. Here, we do not find any impact of our treatment on

[8] In an additional specification, we dropped the ESCS index, which is an aggregated index of the socioeconomic background of the students, and we controlled separately for the labour market situation and the level of education of the mother and the father. Our results did not change significantly. These results are available upon request.

students' performance. The coefficient switches from positive to negative from one specification to the other and it is not statistically significant in any of them. The lack of significance could be partly explained by the fact that we are using as student outcome PISA scores in a subject, which was the focus of the PISA exam only in 2003 and not in 2009, so comparisons are not fully valid.

The different result for Reading and Mathematics could be explained by the fact that Reading was the main focus of the PISA test in 2000 and 2009, whereas Mathematics was the main focus in 2003 but not in 2009. Being the focus of the study means, as stated in the Institutional setup, that two thirds of the study is devoted to this competency whereas the other two thirds are equally divided for the other two areas. This more detailed analysis on one of the competencies allows for disaggregating the students' outcomes by subscales in the chosen competency. In 2009, PISA detailed the levels of student proficiency in various aspects of reading, such as students' ability to access and retrieve, integrate and interpret, and reflect and evaluate the information they obtain through reading. It also examined students' ability to read and understand continuous and non-continuous texts. In contrast, PISA 2009 only computed an aggregated outcome for Mathematics and Science. Thus, the Reading outcome is a more precise estimation than the Mathematics outcome in 2009 whereas in 2012 it will be the other way around.

Nevertheless, the data seem to indicate that something differential has happened in Madrid between 2000 and 2009 with respect to other Spanish regions. A natural hypothesis in this context is that the introduction and publication of the results of standardized exam played a major role in this change. It is very hard to provide a definitive proof with these data, but we can discard some alternative explanations.

Public spending in education per pupil affects to some extent students' outcomes (OECD, 2010). The Spanish Ministry of Education provides data on public spending on education per pupil by regions starting from 2004. During the period 2004-2009, Madrid increased public education spending per pupil by 21%, less than the Spanish average of 33%. More importantly, Madrid has been the region increasing less the education spending per pupil among all the 17 Spanish regions. So, education expenditure cannot explain the better behavior of Madrid PISA scores.

Spain received a large amount of immigrants between 2000 and 2009 (immigrants went from 2% to 12% of the population over the period, Instituto Nacional de Estadística, INE), and Madrid was a major place of destination (it has about 18% of the immigrants and about 13% of the population). But our data can identify whether the student is immigrant and the number of immigrants vary enough between schools so that their effect is probably captured at the school level. This was also a period of rapid economic growth, which was not identical between regions, but the ESCS index has enough information about this variable at the individual level to properly control for the effect of economic data. Some other factors affect schools more directly. Madrid has a larger number of charter schools than other regions. Madrid has also increased the share of charter schools but this trend has been similar to the rest of Regions, if anything a little bit smaller. In any case since the identity of the schools is observable its effect can be controlled.

The only other important institutional reform in Madrid school in this period, beyond the introduction and publication of external exams, is the introduction of *bilingual schools* in the region, where English is a medium of instruction for at least one third of the school time.[9] Although this is clearly an important reform, it has been implemented only gradually starting from first grade, and the oldest students exposed to the program are now 13 years old. In addition, Anghel, Cabrales and Carro (2012) have not found significant effects of the program in either language or mathematics, and possibly a negative effect on natural and social science (the subjects taught in English).

5. Synthetic control method

In this section, we use the methodology proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010), which applies synthetic control methods to comparative case studies. Their methodology is motivated by the fact that, in comparative case studies, the researcher is usually forced to find similarities between treated and non-treated units using observable characteristics, something that it is often difficult in practice. To solve this problem they propose to construct a combination of units for comparison purposes, since the combination will typically resemble the treated unit much better than any single unit alone.

In our case we construct a combination of Spanish regions that resembles the region of Madrid in terms of various characteristics before the treatment and we observe the evolution of this combination in the absence of treatment. This combination is called a synthetic control group. It is constructed by searching for a weighted combination of the untreated Spanish regions, in terms of various predictor variables, which are averaged over the entire pre-intervention period. According to Abadie et al. (2010) "because the choice of

[9] Students not only study English as a foreign language but also some subjects (at least Science, History and Geography) are taught in English. Spanish and Mathematics are taught only in Spanish.

a synthetic control does not require access to post-intervention outcomes, the synthetic control method allows researchers to decide on study design without knowing how those decisions will affect the conclusions of their studies".

In order to construct the synthetic control group (the synthetic Madrid), we have to aggregate the data at school level and then at region level. The year the CDI standardized exam was launched in the region of Madrid was 20004/05, therefore we have two years of pre-treatment data in PISA (PISA 2000 and 2003). PISA 2006 and PISA 2009 will be our post-treatment period. The synthetic Madrid is constructed as a weighted average of the pool of untreated regions. Our donor pool includes 15 regions.[10] The weights are chosen so that the resulting synthetic Madrid resembles the real Madrid as closely as possible in terms of the values of a set of predictors of students' performance before the introduction of the CDI exam, that is, before the treatment.

We include in the list of predictor variables for calculating the weights the following variables: teacher/student ratio, ESCS school index, proportion of immigrants in the school, proportion of public schools in the region, proportion of private schools in the region, proportion of charter schools in the region, proportion of schools in cities (between 100.000 to about 1 million people), proportion of mixed families (students living with a mother and a guardian, with a father and a guardian or with two guardians), average age in the school, average learning time (in hours/week). All variables are averaged at region level and over the pre-intervention period (2000 and 2003).

Using these predictor variables we construct the synthetic Madrid as the convex combination of regions, which most closely resembles the region of Madrid in the pre-treatment period. Table 5 shows the characteristics of the real Madrid region, the synthetic Madrid region and the donor pool (the average of the 15 regions which form the donor pool), in terms of the control variables. The figures prove that the constructed synthetic Madrid is more similar to the real one, in both Reading and Mathematics, than the simple average of the regions that form the donor pool. In Reading, the student/teacher ratio in the real Madrid is 16.18 and in the synthetic Madrid it is 14.25 (the average of the control group is 14.03). The school average of ESCS index is -0.16 in the real Madrid and -0.18 in the synthetic Madrid (the average of the control group is -0.31). The proportion of public schools in the real Madrid is 50% and in the synthetic Madrid is 50.55%, while in the donor pool is 57.14%. Likewise, when we look at the proportion of private/charter schools we observe proportions more similar between the real and the synthetic Madrid than between

[10] There are 17 regions (including Madrid region) and two autonomous cities (Ceuta and Melilla) in Spain. We had to drop Baleares and Ceuta and Melilla because of missing data, so this leaves us with 15 regions.

the real Madrid and the average of 15 regions. The same applies to the proportion of mixed families and the average learning time. Finally, there is a substantial difference in the percentage of immigrant students between the real Madrid and the synthetic Madrid. Furthermore, we find affinities between the synthetic and the real Madrid in students' PISA outcomes as well.

Table 5: Predictor Means for PISA Student Performance

	READING		**MATHEMATICS**		**READING**	**MATHEMATICS**
	Madrid		Madrid		**Average of 15 regions**	
Predictor Balance	Real	Synthetic	Real	Synthetic		
Student/teacher ratio	16,18	14,25	16,18	13,36	14,03	14,03
School average of ESCS index	-0,16	-0,18	-0,16	-0,20	-0,31	-0,31
% immigrant students in school	5,43	2,82	5,43	3,32	2,77	2,77
% Public schools	50,00	50,55	50,00	49,88	57,14	57,14
% Private schools	10,28	9,90	10,28	11,97	7,16	7,16
% Charter schools	29,17	27,01	29,17	30,58	28,41	28,41
% Schools in cities (between 100.000 to about 1.000.000 people	23,06	38,27	23,06	34,49	38,75	38,75
% Mixed families	2,37	2,34	2,37	2,75	2,11	2,11
Average school age	15,82	15,81	15,82	15,81	15,83	15,83
Average learning time (hours/week)	3,26	3,21	2,96	2,92	3,21	2,86
PISA score 2000	506,80	504,83	490,83	487,36	496,92	482,30
PISA score 2003	490,66	492,84	492,11	496,06	488,63	493,23
PISA score 2006	477,27	477,71	500,11	494,02	471,05	492,25
PISA score 2009	500,86	489,45	494,63	495,60	484,70	490,64

Note: All variables, except PISA scores, are averaged at region level for the pre-treatment period (2000 and 2003). Learning time in Reading for average of 15 regions is for 2000 (there is no data for 2003).

Table 6 displays the weights of the 15 regions from the donor pool in the synthetic Madrid. It shows that the students' performance in the region of Madrid is best approximated by a combination of Aragón, Asturias, Canary Islands, Cataluña, La Rioja, Murcia y País Vasco for Reading and Aragón, Asturias, Canary Islands, Cataluña, Murcia y País Vasco for Mathematics. The rest of the regions in the donor pool are assigned zero weights.

Table 6: Region weights in the Synthetic Madrid

Region	**Weight Reading**	**Weight Mathematics**
Andalucia	0	0
Aragón	0,226	0,093
Asturias	0,144	0,229
Canary Islands	0,047	0,038
Cantabria	0	0
Castilla - La Mancha	0	0

Castilla-León	0	0
Cataluña	0,189	0,26
Extremadura	0	0
Galicia	0	0
La Rioja	0,089	0
Murcia	0,248	0,098
Navarra	0	0
País Vasco	0,056	0,281
Valencia	0	0

The next two graphs show the evolution of the real Madrid and the synthetic Madrid in 2000 and 2003 (the pre-intervention years) and in 2006 and 2009 (the post-intervention years), separately for Reading and for Mathematics.

Figure 1: The evolution of the real Madrid and the synthetic Madrid in Reading in 2000, 2003, 2006 and 2009

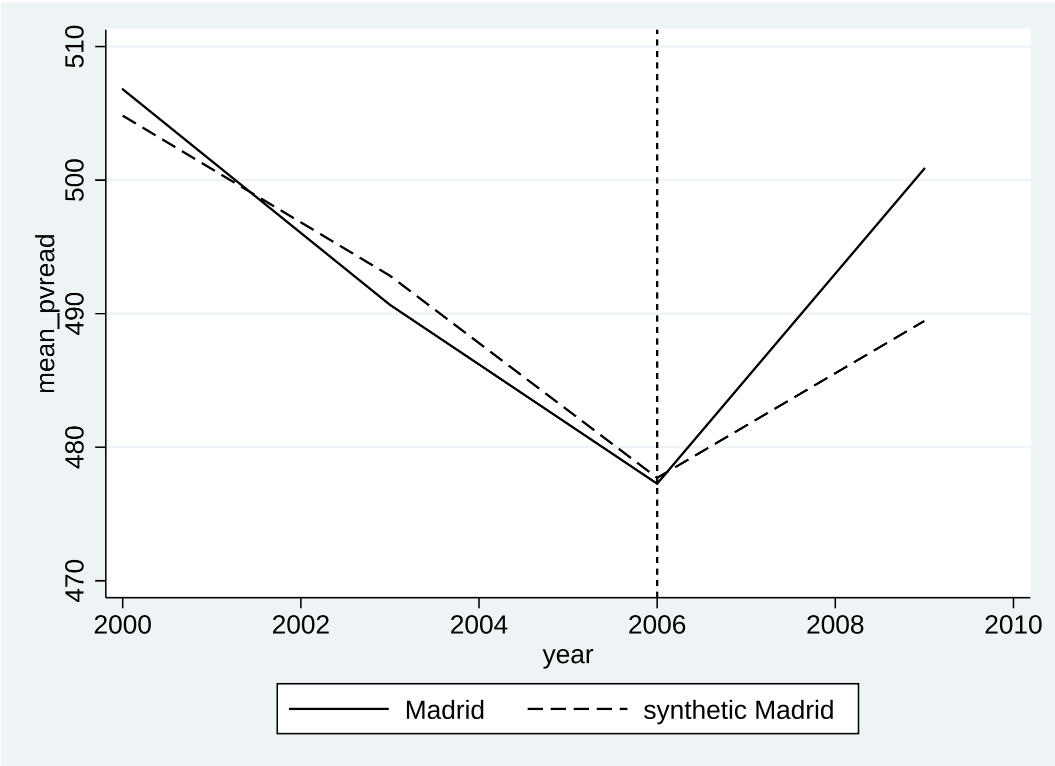
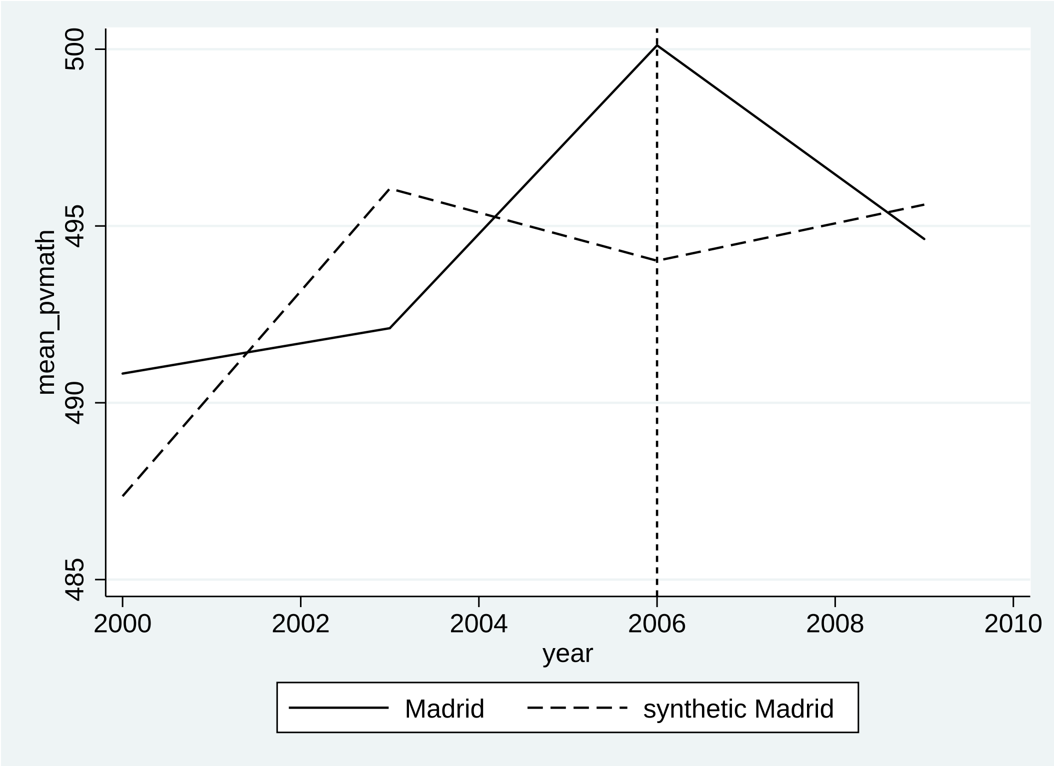


Figure 2: The evolution of the real Madrid and the synthetic Madrid in Mathematics in 2000, 2003, 2006 and 2009



For Reading, the graph shows that the synthetic Madrid approximates very well the evolution of the real Madrid in the pre-treatment period. After the treatment, which we take to be the introduction of the CDI standardized exam in the region of Madrid, in 2006 and 2009, the evolution of the synthetic control group is different than the one of the treatment group. In particular we see that, even if both real Madrid and synthetic Madrid experience an increasing trend after 2006, synthetic Madrid is doing worse that the real Madrid: in 2009, the difference in performance is 11,41 PISA points in favour of real Madrid. This difference could be attributed to the introduction of a standardized exam in the region of Madrid, at least with the information that we can observe. This confirms, even quantitatively, the results we obtained previously with the diff-in-diff methodology, where we found that controlling for school characteristics, the region of Madrid improved its performance relative to other regions of Spain in the period between 2000 and 2009, between 14 and 17 PISA points. The flagship education publication of the OECD, Education at a Glance, arrives to a similar conclusion in the latest 2012 edition: “students in school systems that use standards-based external examinations score 16 points higher, on average across OECD countries, than students in school systems that do not use these examinations (Education at a Glance, 2012, page 527). Our estimation is a little bit lower than the range found in the literature by Hanushek and Woessmann (2011) of 20% to 40% of the standard deviation (20 to 40 points in PISA).

For Mathematics, however, the synthetic control group methodology does not work so well. The synthetic Madrid does not approximate very well the evolution of the real Madrid in 2000 and 2003, the pre-treatment period. In the post-treatment period, the synthetic Madrid is performing slightly better than the real Madrid. Nevertheless, the diff-in-diff estimation showed no statistical impact in Mathematics.

We are aware of the limitations of our data in performing the estimation by using synthetic control methods. One of them is that since the PISA study started in 2000 and it is carried out each three years, we only have two years of pre-intervention data (2000 and 2003), which complicates the calculation of the region weights for the synthetic control group. The result in Mathematics, where the synthetic Madrid is not so similar to the real Madrid in the years before the treatment can be partly explained by this fact.

6. Conclusions

This paper attempts to identify whether the implementation and publication of the results of external and standardized tests could have any impact on the performance of students. We use the fact that in the region of Madrid a standardized exam was first administered (and its results published) in 2004/05 to all 6th grade primary students, while in the other regions of Spain, no such exam existed. Using a diff-in-diff strategy we find a positive effect in Reading of the order of 14 to 17 PISA points. The synthetic control method yields an effect that is very close in quantitative terms. Our results are in line with previous research in the area, but our study provides one important innovation, since the external exams in Madrid have no consequences for the students, so the effect has to come from the impact on teachers and school principals.

We have identified a possible effect in language, but not in mathematics. This is slightly surprising since many educational programs have observed effects that are larger in mathematics than in language (see e.g. Abdulkadiroglu et al., 2011). A possible explanation may come from the different emphasis of the curricula of primary school education degrees in Spain with respect to other countries, but this question deserves a more thorough investigation, which we defer to further research.

References

Abadie, Alberto and Javier Gardeazabal (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country", *The American Economic Review,* 93 (1), 113-132

Abadie, Alberto, Alexis Diamond and Jens Hainmueller (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program", *Journal of the American Statistical Association*, 105 (490), 493-505

Abdulkadiroglu, Attila, Joshua Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak (2011), "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *The Quarterly Journal of Economics*, vol. 126, pp. 699–748.

Anghel, Brindusa and Antonio Cabrales (2010), "The Determinants of Success in Primary Education in Spain," Working Paper FEDEA, No. 2010-20.

Anghel, Brindusa, Antonio Cabrales and Jesús M. Carro (2012), "Evaluating a bilingual education program in Spain: the impact beyond foreign language learning", CEPR Working Paper, No. 8995.

Bishop, John H. (1997), "The Effect of National Standards and Curriculum-Based Examinations on Achievement", *American Economic Review,* vol. 87 (2), pp. 260-264.

Bishop John H. (2006), "Drinking from the fountain of knowledge: student incentive to study and learn," In: Hanushek EA, Welch F (eds) *Handbook of the economics of education*. North-Holland, Amsterdam

Bishop, John H., Ferran Mane, Michael Bishop and Joan Moriarty (2001), "The Role of End-of-Course Exams and Minimum Competency Exams in Standards-Based Reforms," *Brookings Papers on Education Policy* vol. 4, pp. 267-345.

Hanushek, Eric and Ludger Woessmann (2011): "The Economics of International Differences in Educational Achievement", in Hanushek, E, Machin, S. and Woessmann, L. (Eds.) *Handbook of the Economics of Education*, vol. 3, pp. 89-200, Amsterdam: North Holland

Jürges, Hendrik, Kerstin Schneider and Felix Büchel (2005), "The Effect Of Central Exit Examinations On Student Achievement: Quasi-Experimental Evidence From TIMSS Germany", *Journal of the European Economic Association*, vol. 3(5), pp. 1134-1155, 09.

OECD (2009), "PISA Data Analysis Manual: SPSS Second Edition".

OECD (2010), "PISA 2009 Results: What Makes a School Successful?" Volume IV.

Wößmann, Ludger (2010), "Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries", *Jahrbücher für Nationalökonomie und Statistik* vol. 230 (2), pp. 234–270