APPLIED ECONOMICS, 2015
http://dx.doi.org/10.1080/00036846.2015.1111989

The impact of immigrant concentration in schools on grade retention in Spain: a difference-in-differences approach

Francisco Pedraja-Chaparro[a], Daniel Santín[b] and Rosa Simancas[a]

[a]Department of Economics, University of Extremadura, Badajoz, Spain; [b]Department of Applied Economics VI, Complutense University of Madrid, Pozuelo de Alarcón, Madrid, Spain

ABSTRACT
Since the late 1990s, Spain has played host to a sizeable flow of immigrants who have been absorbed into the compulsory stage of the education system. In this article, our aim is to assess the impact of that exogenous increase in the number of immigrant students from 2003 to 2009 on grade retention using Spanish data from PISA 2003 and 2009. For this purpose, we use the difference-in-differences method as a dose treatment capable of detecting whether the immigrant concentration has had a significant effect on student performance. Our results evidenced that their arrival does not on average decrease school promotion rates with respect to 2003 and is even beneficial to native students. However, although the concentration of immigrant students at the same school does have a negative impact on immigrant students generating more grade retention, native students are unaffected until concentrations of immigrant students is above 15%.

KEYWORDS
Difference in differences; education; immigration; PISA

JEL CLASSIFICATION
I21; H41

I. Introduction

There has been a remarkable increase in the foreign population in Spain over the 2000s, with a constantly growing inflow that accounts for almost one third of the total immigrants received by the OECD (Cebrián et al. 2010). This was the result of the expansion of the Spanish economy, motivated largely by the construction sector boom. These immigration rates have slowed down since 2009 and even declined slightly in absolute terms between 2010 and 2012, possibly due to the economic crisis (Sánchez 2013). Throughout this period, there has been a significant change in the composition of the immigrant population according to their countries of origin. In the early days, most immigrants came mainly from Latin America, whereas the percentage of the immigrant population from other European countries, mainly European Union nonmembers, increased notably towards the end of this period (Rojas and Sánchez 2011).

A direct consequence of this phenomenon is the higher proportion of immigrant students in the Spanish education system, rising from 1.5% in 2000 to 9.5% in 2011 with a 9.81% peak in 2009. Table 1 shows immigration figures in Spain from 2000 to 2011 and the evolution of the proportion of immigrant students in the Spanish education system.

In most countries, immigrant students have lower educational outcomes, higher dropout rates and lower levels of noncompulsory education than native students (Driessen 2001; Schnepf 2008; Murat and Frederic 2015). Studies focusing on average differences in educational outcomes between immigrant and native students from traditionally immigrant-receiving countries like Germany provide evidence that immigrant students are not able to definitively close the educational gap between themselves and their native classmates (Frick and Wagner 2001; Ammermueller, 2007). In some other countries like Belgium and Canada, however, where native students continue to outperform their immigrant peers, the performance gap has narrowed despite the rising percentage of immigrants (Entorf and Minoiu 2005; OECD 2011). Nevertheless, there are nonconclusive evidences about the impact of high rates of immigrant pupils on the achievement of natives. On the one hand, Brunello and Rocco (2013), using cross-country data from 19 different countries, and Contini (2013), analysing Italian

CONTACT Francisco Pedraja-Chaparro pedraja@unex.es

Table 1. Data about immigrant population in Spain.

Year	Immigrant Population	% Total Population	% Immigrant Students in the Education System
2000	923 879	2.3	1.5
2001	1 370 657	3.3	2.0
2002	1 977 946	4.7	2.9
2003	2 664 168	6.2	4.4
2004	3 034 326	7.0	5.7
2005	3 730 610	5.5	6.5
2006	4 144 166	9.3	7.4
2007	4 519 554	10.0	8.4
2008	5 220 600	11.3	9.4
2009	5 598 691	12.0	9.8
2010	5 747 734	12.2	9.7
2011	5 730 067	12.2	9.5

Source: Author's calculations using data from the municipal register (National Institute of Statistics).

educational system, highlight that high proportions of immigrant students affect negatively the natives' learning outcomes, although the size of this effect is relatively small and weak. On the other hand, Ohinata and Van Ours (2013) and Geay, McNally, and Telhaj (2013) using data from primary education in Netherlands and England, respectively, do not find any negative impact of concentration of immigrant students on the performance of native ones. The latter even detects a positive effect for Maths achievement of native pupils from England, but only when the impact of the increase of immigrant students is analysed within Catholic schools. The authors suppose that this positive impact is due to the fact that these immigrant students are from Eastern European families whose parents are highly educated. In Spain, recent articles have studied this phenomenon using different approaches: Calero and Waisgrais (2009) and Calero, Choi, and Waisgrais (2009) compare the educational performance of immigrant students and their peers using multilevel regression techniques, concluding that the determinants of educational achievement affect native and immigrant students differently. Zinovyeva, Felgueroso, and Vazquez (2014) perform Oaxaca-Blinder decomposition in order to analyse the educational gap between natives and immigrants and find that around half of this gap can be attributed to socioeconomic and family factors. Finally, Salinas and Santín (2012) employ a switching regression model to calculate the impact of immigration on the educational outcomes controlling for school type. They show that immigrant students have a higher probability of attending public schools and that the negative effect on native students produced by the concentration of immigrants is bigger in public schools than in private government-dependent schools.

Another relevant issue in Spain is the high rates of grade retention (around 30% of students), which is a warning sign of school failure and a good predictor of school dropouts. Several studies support the hypothesis that repeating a grade is often the main predictor of school failure (Roderick 1994; Jimerson, Anderson, and Whipple 2002; Benito 2007). This has led us to study the effect of immigration from another perspective. We consider whether or not the increase in immigrant students recent years has had repercussions on grade retention rates particularly for native students. A similar approach is followed by Cristia, Czerwonko, and Garofalo (2014) in order to test whether increased technology access in schools affect retention rates.

This article uses an impact evaluation approach to study how the increase of the proportion of immigrant students in some schools can affect grade retention rates. For this purpose, we estimate the impact of the exogenous increase of immigrant students[1] in Spain from 2003 to 2009 using a Difference-in-Differences approach (DiD).[2] Using this technique, we can determine whether the concentration of immigrants has a significant effect on student performance by comparing the percentages of students studying in the proper grade by age. A similar approach could be broadly applied to several developed countries, since nowadays the integration of immigrants into society becomes a main concern around the world. The same design could be also used to the so-called new immigration countries, those which transformed from immigrant-sending countries to immigrant-receiving ones. These countries such as Portugal, Italy or Greece have increased considerably their foreign population in the past

[1]*Native students* are students born in the country of assessment or who have at least one parent who was born in that country. *Immigrant students* are students who are foreign-born and whose parents are also foreign-born or students who were born in the country of assessment but whose parents were not (OECD 2010); i.e. without differentiating between first-generation immigrants and second-generation immigrants. We include them in the same general category (*Immigrant students*) due to the fact that most immigrant students evaluated in PISA 2003 were second-generation immigrants and the sample of first-generation immigrants would not be representative.

[2]We are aware that the residential choices made by immigrants, as well as their school district choices are nonrandom. However, it is a fact that immigrant population tends to set in lower socioeconomic areas, and these areas do not change between the two periods analysed, as neither do the schools.

decades. In the same way, the traditionally immigrant-receiving countries, such as Germany, France or the UK, could test whether, even being in an advanced stage of the immigration process, this situation keeps having some impact on the educational achievement of the destination country.

The research reported here makes two contributions. First, we apply the DiD method to analyse the possible relationship between the increase of foreign students and grade retention rates. The idea behind this approach is that the treatment types could differ in some situations, depending in this case on the concentration of immigrants. On this ground, the treatment will be referred to as a *dose* treatment. Second, instead of applying this methodology to longitudinal data as is common practice in the previous literature, we use data from consecutive cross sections OECD PISA reports for Spain. This approach of using data from different waves of PISA is similar to the one followed by Zinovyeva, Felgueroso, and Vazquez (2014) who analyse the gap between immigrant student's achievement with respect to native students in Spain through Oaxaca-Blinder decompositions.

The article is structured as follows. Section II presents and justifies the applied methodology. In Section III, we describe the data set used and the selected variables included in the empirical analysis. Section IV reports the results. We conclude in Section V by discussing the implications of our findings for public policy.

II. Methodology

The goal of our research is to analyse the impact of the growth in immigrant students experienced by Spain over the last 10 years on the average grade retention rates per school. According to the theory of impact evaluation, having foreign students enrolled at the school would be the *treatment*, and schools with immigrants would be the *treated* schools. Note, however, that this is a *dose treatment*, so we are not simply looking for the average effect of there being or not being foreign students at the school, but the effect of their concentration on the treated schools. Therefore, we have two groups. One group is composed of schools hosting the immigrants, considered as the *treated group*. These schools will have also received different treatments because the concentration of immigrants varies over time. The other group includes schools not hosting immigrants, known as the *nontreated* or *control group*.

The rate of nonrepeater students (who are in the correct grade) from 2003 to 2009 at the control schools will vary due to a number of possibly unknown factors. The variation of this rate at the treatment schools will be due to the same factors plus the variation in the component we are trying to evaluate, i.e. the arrival of immigrants. In order to estimate the impact of the exogenous increase in the number of immigrants, we use the DiD technique by means of which we can isolate the effect of immigrant arrival from the unknown factors. Although this technique requires panel data, it can also be estimated using cross-sectional databases, provided that they can be guaranteed to be consistently representative (Khandker, Koolwal, and Samad 2010), and the samples are selected according to the same procedure throughout (Meyer 1995). In this case, consecutive PISA reports (OECD 2004, 2010) satisfy these requirements.

The DiD method calculates the average difference in outcomes separately for treatment and nontreatment groups over the period. Then, after taking an additional difference between the average changes in outcomes for these two groups, it is possible to identify the difference-in-differences impact, i.e. the estimated impact of the assessed issue. For our empirical educational model, let Y_t^{T} and Y_t^{C} denote the mean percentages of students in the proper grade for their age at treated and control schools, respectively, and t a dummy variable that can take two values: 2003 and 2009. The classical DiD technique estimates the average impact as follows:

$$\mathrm{DD} = E\left(Y_{2009}^{\mathrm{T}} - Y_{2003}^{\mathrm{T}}\right) - E\left(Y_{2009}^{\mathrm{C}} - Y_{2003}^{\mathrm{C}}\right) \quad (1)$$

Note that if the treatment group differs from the control group in terms of observed and unobserved characteristics in addition to treatment, we need to assume that the differences between the two groups are time-invariant in order to obtain an unbiased difference-in-differences estimator. The DiD estimator can be solved using a regression. On the basis of the discussion in Ravallion (2008), the estimating equation would be as Equation 2:

$$Y_t^D = \alpha + \beta Dt + \rho D + \gamma t + \varepsilon \quad (2)$$

where D is the treatment variable, t is the time dummy variable and the coefficient of the interaction of D and t, β represents the estimated impact of the treatment on outcome Y:

$$D = \begin{cases} 1 \text{ if it belongs to the treatment group} \\ 0 \text{ if it belongs to the control group} \end{cases}$$

$$t = \begin{cases} 1 \text{ if year} = 2009 \\ 0 \text{ if year} = 2003 \end{cases}$$

The coefficient of the interaction β indicates whether or not the increase in immigrant students has a significant impact on the dependent variable and how much impact it has. In addition to the interaction term, the variables time (t) and treatment (D) are also included in order to detect any isolated effects due to the time or to group membership.

As mentioned at the beginning of this section, we are not only interested in measuring the average effect of immigrant students on educational performance, but also the impact of their concentration. For this reason, we include what we call a *dose treatment* in our research, and these *doses* are the percentages of immigrants at each school belonging to the treated group, represented by the variable *Immig*.[3] Although dose treatments usually consider finite numbers of treatment levels (i.e. a discrete variable such as different cash transfer sums), this approach can also be applied to continuous treatments (Abadie 2005), as is in this case. The explanatory variable *Immig* is added to a saturated model combined with time, treatment and the interaction of both variables. However, the saturated model cannot be estimated because of its perfect multicollinearity.

Since we are only interested in the term that contains the treatment dose ($\delta_2 \text{Immig}Dt$), the equation we finally estimate is as follows:

$$Y_t^D = \alpha + \beta Dt + \rho D + \gamma t + \delta \text{Immig}Dt + \varepsilon \qquad (3)$$

The DiD estimator is now the result of adding two terms: the interaction coefficient β and the effect that contains the percentage of immigrants δImmig.

According to the specification, first we have the treatment (having immigrant students at school), which leads to the average impact of the increase of immigrant students between 2003 and 2009; and second, we include the dosage (percentage of immigrant students) which corrects the average effect since it allows that the concentration of immigrant has a different impact over the average one.[4]

We can summarize our strategy as follows. In the first period, we have two groups: schools with and without immigrants. Across the two periods, we assume that immigrant students join the education system and enrol in the schools. This is equivalent to increasing the *dose* of immigrants in the education system, and we are interested in analysing the impact of this increase on grade retention. At the end of this period, we again have schools with no immigrant population (the control group) and schools with a higher mean percentage of immigrants (the treated group), although this mean is not uniformly distributed across schools. This implies that the dose received by each treated school is different.[5]

It is noteworthy that a basic assumption behind this technique is that the remaining covariates (X), which could affect both the treated and the control groups, must be unchanged over time. If this is not a valid assumption, the regression analysis should control those covariates in order to ensure a correct estimation as follows:

$$Y_t^D = \alpha + \beta Dt + \rho D + \gamma t + \delta \text{Immig}Dt + \eta X + \varepsilon \qquad (4)$$

In this case, the regressions include five control variables. They are described in the following section. Furthermore, the trends of the treatment group and the control group are assumed to be equal in the absence of treatment, although this assumption cannot be tested. However, we performed a *placebo* test in order to check the validity of the DiD method. This test involves performing an additional DiD estimation using a *fake* treatment group (i.e. comparing two control groups) or a *fake* outcome (Gertler et al. 2011). Because of the type of database, we chose the second option, using the *average percentage of girls per school* as our fake dependent

[3]This idea is closely related to the approach developed by Abadie and Dermisi (2008).

[4]The same estimates can be done taking out the main treatment factor, letting alone the treatment intensity factor without significant changes in the results.

[5]It may be possible that some schools in the control group in the first period were classified in the treatment group during the second period, but the opposite is unlikely. This fact guarantees that control group samples are similar in terms of composition in both periods.

variable uncorrelated with the treatment, as is also performed by Felfe, Nollenberger, and Rodríguez-Planas (2015).

Finally, the results section includes a simulation analysis of how the average promotion rates per school vary depending on the percentage of immigrant students enrolled in order to clarify our estimations.

III. Data and variables

The PISA report

The data set used for the research comes from the PISA (*Programme for International Student Assessment*) survey, designed by the OECD in 1990s as a comparative, international, regular and continuous study on certain educational characteristics and skills of students worldwide (Turner 2006). The PISA target population is composed of students who are aged between 15 and 16 years old at the time of the assessment, all of whom are born in the same year and who have completed at least six years of formal schooling. PISA measures their performance in math, reading and science. It also collects information about students' personal background and schools environment, for which purpose two questionnaires are administered, one addressed to school principals and another to students.[6] These surveys have taken place every three years since the year 2000 focusing on one of the above three areas each time.

An important aspect that is to be taken into account in an empirical analysis using PISA data is that the data are gathered by means of a two-stage sampling procedure. First, a sample of schools is selected in every country from the full list of schools containing the total student population. Then, a sample of 35 students is randomly selected within each school. As a result, statistical analyses have to consider sampling weights in order to ensure that sampled students adequately represent the analysed total population (Rutkowski et al. 2010).[7]

Sample, variables and the identification strategy

Although the DiD method usually uses panel data, repeated cross-sectional data from the same areas has also been used in the literature (Chaudhury and Parajuli 2010; Felfe, Nollenberger, and Rodríguez-Planas 2015[8]). We use data from two different waves, 2003 and 2009, which provide information useful for interpreting average results concerning the 2002/03 and 2008/09 academic years. The chosen unit of analysis is the school, and therefore, the data is aggregated at school level. PISA samples are composed of different school types that can be divided into three groups according to their ownership: public (government managed and funded schools), private (privately managed and funded schools) and private government dependent (privately managed and government funded schools). In our research, we focus on schools that are comparable in terms of public funding and also share the same admission criteria,[9] i.e. public and private government-dependent schools. The sample is composed of 336 schools (199 public schools and 137 private government-dependent schools) in 2003 and 806 schools (512 public schools and 294 private government-dependent schools) in 2009.[10]

Regarding the variables, we use the *percentage of students who are in their correct grade* (without repeating any year) and the *percentage of native*

[6]Parents complete a third questionnaire. However, this information is only available for a limited number of countries and, unfortunately, Spain is not one of them.

[7]These weights include adjustments for nonresponse by some schools and students within schools and weight cutting to prevent a small set of schools or students having undue influences. These processes are based on intensive calculation methods, known as resampling methods, which consist of taking multiple samples from the original sample. Specifically, PISA uses the Balanced Repeated Replication (BRR) with 80 replicates. For an extensive description of this procedure, see (OECD 2005; OECD 2009a).

[8]This study specifically applies the same methodological approach to the same database that let us (different waves of PISA) to test the impact of a substantially public childcare expansion in Spain.

[9]Public-funded schools cannot reject immigrant students that ask for a position in the school. This fact prevent the model of potential bias results occurring if we had that schools with no immigrants were the result of selection. On the other hand, note that immigrant students attending private schools are a minority that can afford an expensive education, and they are not supposed to generate any educational problem.

[10]The difference in sample size between the two periods is due to the fact that PISA 2009 covered more regions with an extended sample than PISA 2003 (14 regions in 2009 and 3 regions in 2003). However, both samples can be used to obtain general conclusions for Spain due to the fact that both PISA 2003 and PISA 2003 are nationally representative.

students who are in their correct grade as dependent variables.[11] Since PISA assesses 15-year-old students, we consider that 4th-grade ESO students (the so-called *Enseñanza Secundaria Obligatoria*, i.e. compulsory secondary education in the Spanish system, equivalent to 10th grade on the international scale) are in their correct year.[12] We differentiate between these two dependent variables in order to distinguish how the concentration of immigrant students in schools affects grade retention and native grade retention, in particular.

In our analysis, the treated schools are schools that have immigrant students. As the distribution of immigrant students is not uniform across the education system, the concentration of these students differs from one school to another. As we described in the methodology section, the aim of introducing this issue in our econometric models, we consider a *dose treatment*. In this way, we include the *percentage of immigrants* (*Immig*) in the base model (2), defined as the ratio between immigrant students and the total number of students sampled by school in order to capture the potential effects of a higher presence of immigrants in schools (3).

The school distribution by control and treated groups, and the different treatment doses are shown in Tables 2 and 3, respectively.

Table 2. School distribution by groups.

	2003		2009	
	Schools	%	Schools	%
Control Schools	154	45.8	168	20.8
Treated Schools	182	54.2	638	79.2
N	336	100	806	100

Source: Author's calculations using data from PISA (OECD 2004, 2010).

Table 3. Different treatment doses within treated schools.

	2003		2009	
Treated Schools: Immig Dose	Schools	%	Schools	%
<5%	81	44.50	136	21.32
5%–10%	54	29.67	161	25.24
10%–15%	27	14.83	119	18.65
15%–20%	7	3.85	79	12.38
20%–25%	7	3.85	49	7.68
>25%	6	3.30	94	14.73
Total	182	100	638	100

Source: Author's calculations using data from PISA (OECD 2004, 2010).

From Table 2 we conclude that the percentage of schools with immigrants grew significantly from 2003 (54.17% of total) to 2009 (79.17% of total). Additionally, Table 3 shows that around 11% of schools had an immigrant student population of more than 15% in 2003, whereas this percentage multiplied by more than three in 2009 reaching 34.79%.

Moreover, as we explained above, we select a set of control variables to be introduced in the model (names in brackets denote variable names in the results tables).

Concerning parental background we included, the *Index of parental occupational status* (Parental Occupation) represents the index of highest occupational status of parents according to the *International Socio-Economic Index of Occupational Status* (ISEI, Ganzeboom, De Graaf, and Treiman 1992). We built a variable that represents the average value of this index for each school. We assume that the higher the average parental occupational status, the greater their income, whereby students enrolled at this school will have higher average socioeconomic status and the *Parental educational level* (Parental Education), an index of highest educational level of parents in years of education according to the *International Standard Classification of Education* (ISCED, OECD 1999). Again, we construct a variable that represents the average value of this index for each school.

As regards school characteristics, we selected the *Type of School* (School Type) as a *Dummy* variable that takes value 1 if the school is a private government-dependent school and 0 for a public school and the *Quality of school resources* (School Resources) which is a continuous variable based on the school principal's responses to seven questions available from PISA 2003 and PISA 2009 databases related to the availability of computers for educational purposes, educational software, calculators, books, audiovisual resources and laboratory equipment.

In order to control for the school location, we introduced four dummy variables related to the town population: *Village, Small town, City* (taken

[11]Grade retention is chosen as dependent variable instead of student test scores due to the fact that the values of test scores in PISA are rescaled each wave (OECD average equals 500 and SD 100). Therefore, it is impossible to make comparisons of one country performance over time because normalization avoids concluding if the output is really increasing or decreasing over time. For more detail, see PISA 2006 Technical Report (OECD 2009b).

[12]We can do this assumption due to the fact that in Spain there are legal constraints preventing parents from choosing their children's enrolment cohort. Parents cannot postpone their children's entry to the first year of primary school with the aim of their children being more mature and performing better at school.

Table 4. Descriptive statistics.

	2003				2009			
	Control		Treated		Control		Treated	
Schools	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Dependent variables								
% Students in the correct year	0.7342	0.1666	0.7073	0.1847	0.7310	0.19482	0.6508	0.1764
% Native students in the correct year	0.7342	0.1666	0.6634	0.1852	0.7310	0.19482	0.5995	0.1869
Independent variables								
% Immigrant students (Immig)	0.0000	0.0000	0.0835	0.1044	0.0000	0.0000	0.1394	0.1273
Parental Occupation	43.2738	8.4051	43.7276	7.7022	47.4312	10.1658	44.6906	7.6220
Parental Education	11.4560	1.6249	11.2753	1.6100	12.6508	2.0495	12.1989	1.6387
School Type	0.4400	0.4980	0.3800	0.4870	0.5200	0.5010	0.3200	0.4680
School Resources	−0.0393	0.9982	−0.0932	1.0074	0.0332	0.7855	−0.0156	0.8472

Source: Author's calculations using data from PISA (OECD 2004, 2010).

Table 5. Distribution of schools within the different population sizes.

	2003		2009	
Regions	Control	Treated	Control	Treated
Village (Pop < 15 000)	50	54	71	202
Small Town (Pop. 15 000–100 000)	52	53	42	217
City (Pop. 100 000–1 000 000)	49	66	53	198
Large City (Pop. > 1 000 000)	3	9	2	21

Source: Author's calculation using data from PISA (OECD 2004, 2010).

as the baseline category) and *Large City*. Each dummy variable takes value 1 if the school is located in a town with an amount of population within the bounds specified (in Table 5, it can be checked the different bounds for every dummy variable). Tables 4 and 5 report the main descriptive statistics for the variables considered in our analysis and the distribution of control and treatment schools within the different population sizes.

It is well-known in the literature the existence of two factors that must be taken into account when the educational achievement of immigrant students is being analysed: their country of origin and whether they are familiar with the language spoken in the country of destination. However, these variables could not be included in the current empirical analysis as control variables due to the lack of information in the PISA 2003 and PISA 2009 surveys, at least in the Spanish case.

Table 4 and 5 report the main descriptive statistics for the variables considered in our analysis and the distribution of control and treatment schools within the different population sizes.

According to Table 4, regarding the dependent variables, it can be seen that both the *percentage of students in the correct year* and the *percentage of native students in the correct year* in the control group are quite similar comparing the year 2003 to the year 2009. On the other hand, the rates in both dependent variables have experienced a decrease between the two periods of time assessed in the treated group. This decline is higher in case of the second dependent variable, what suggests us that the percentage of immigrant students in the correct year (variable not presented in Table 4, but can be obtained by subtracting the second dependent variable from the former one) has slightly increased. This result might be due to higher amount of immigrant students in the year 2009 compared to year 2003 and the improvement of Spanish educational system in terms of integration.

In relation to the independent variables, our *dosage* (percentage of immigrant students enrolled in the school) has considerably risen between the first and the second year. Comparing the evolution of the indexes of *parental occupation status* and *parental education level* in both groups between the year 2003 and year 2009, it can be said that these variables increased over the two years both in control group and the treatment group, suggesting that the average levels of parental occupation and parental education have improved throughout this period of time in Spain coinciding with the economic boom. It can be claimed the same for the school resources index. Focusing on the school type variable, it can be noted that there is a rise of private government dependent schools without immigrant students during the considered time (in relative terms). In our opinion, it is because the combination of two different factors. First, the number of private government schools who provide compulsory secondary education increased in greater extent than the number of public schools between the two years. Second, it can suggest the existence of some degree of segregation of immigrant students, and this intuition can be, somehow, right due to the fact that private government-dependent schools often locate in larger cities and neighbourhoods and, on average, in higher socioeconomic status areas within the cities, where some immigrant families can be found living around but not the most of them.

IV. Analysis of results

Results

This section presents the results for the models described in the methodology. Specifically, we estimate three different models for each dependent variable: *percentage of students in their correct grade* (Students) and *percentage of native students in their correct grade* (NStudents). Model 1 is the basic difference-in-differences model estimation (2). Model 2 is equivalent to the basic model plus the treatment dose (3) captured through the percentage of immigrants at the school combined with the interaction term (δImmig). Finally, Model 3 estimates Equation 4 as an extension of Model 2, in which control variables are also introduced in order to single out the net effect of treatment. By including these variables, we can test whether or not they have a separate effect on the outcome.

Table 6 reports the model estimation parameters, showing variable coefficients, standard errors and statistical significance in each column. At this point, all effects will be quantified on the average percentage of students who are in the correct grade for their age and, therefore, have not repeated any year.

First, regarding estimates of the *percentage of students in their correct grade* (Model 1) shows that, taken separately, neither the time variable nor group membership has a significant effect on the dependent variable. With respect to the coefficient associated with the interaction term *(β)*, i.e. the difference-in-differences estimator, we observe no significant difference between treated (schools with immigrants enrolled) and control group (schools without immigrants enrolled) throughout the evaluated period. The information provided by the interaction term is the average effect of an increase of immigrants. Thus, given that PISA evaluated schools have few immigrants on average, it is reasonable to assume that, on average, promotion rates at schools with an average number (few) of foreign students do not decrease significantly compared to 2003 with respect to control schools. This result appears to suggest that schools with low mean values have adapted well to this new situation (slight increase of immigrant student enrolment). The addition of the dose treatment in Model 2 discloses similar results related to the above variables. However, the coefficient associated with the interaction term by

Table 6. Difference-in-differences estimations for all students.

Dependent Variable	Model 1		Model 2		Model 3	
All Students	Coeff.	$p > \|t\|$	Coeff.	$p > \|t\|$	Coeff.	$p > \|t\|$
Constant	0.6924 (0.0280)	0.000	0.6924 (0.0281)	0.000	0.0961 (0.0679)	0.158
Year (*t*)	−0.0579 (0.0462)	0.211	−0.0579 (0.0462)	0.211	−0.0992 (0.0388)	0.011
Treatment (T)	0.0003 (0.0342)	0.992	0.0003 (0.0342)	0.992	−0.0067 (0.0262)	0.798
Interaction	0.0026 (0.0510)	0.960	0.0767 (0.0519)	0.140	0.0645 (0.0397)	0.104
Immig (interact)			−0.5499 (0.0705)	0.000	−0.3235 (0.0658)	0.000
Parental Occupation					0.0067 (0.0016)	0.000
Parental Education					0.0180 (0.0077)	0.019
School Type					0.0806 (0.0208)	0.000
School Resources					0.0134 (0.0088)	0.128
Village					0.0226 (0.0196)	0.250
Small Town					0.0176 (0.0193)	0.362
Large City					−0.0267 (0.0249)	0.286

Source: Author's calculations using data from PISA (OECD 2004, 2010).
Note: SEs are presented in parentheses.

the percentage of immigrants ***(δ)***, i.e. the difference-in-differences *dose* estimator turns out to be statistically significant and is negatively related to the dependent variable. This implies that the concentration of immigrant students has a negative impact on grade retention for all students (immigrant and native students) with respect to the control group.

Model 3 parameters illustrated in Table 6 can be interpreted similarly. The only notable difference is that the effect of immigrant concentration persists and is significant, albeit to a lower extent, despite control based on the variables related to school type, school resources, school location and school average socioeconomic status, through indexes that represent the level of parental education and parental occupation. With respect to the control variables introduced in the model, variables representing the educational level and occupational status of parents and the type of school are statistically significant.

Table 7 illustrates the three model estimation parameters for the percentage of native students in their correct grade only.

According to Table 7, the estimation of the *percentage of native students in their correct grade* (dependent variable) shows only one relevant difference with respect to the previous model. In this case, the last two models report a statistically significant interaction coefficient (β) with a positive correlation with the dependent variable. Hence, it can be argued that, when the percentage of immigrants enrolled is introduced (treatment dose), native students benefit on average from having a small number of immigrant students in the classroom. We believe that this effect may be due to the fact that immigrants are more susceptible to suffer grade retention. It is worth to note that this result could also be due to, for instance, the improvement of noncognitive skills of native students because of sharing classroom with their immigrant peers. However, it would be a topic of discussion for a different paper, and it is beyond the scope of the current one.[13] Nevertheless, this slight advantage is offset and, finally, even cancelled out by the *dose* coefficient.

Simulation

To clarify the above results, Table 8 is a simulation of how the average promotion rates vary in schools based on the percentage of enrolled immigrant pupils.[14] Any

Table 7. Difference-in-differences estimations for native students.

Dependent Variable	Model 1		Model 2		Model 3	
Native students	Coeff.	$p > \|t\|$	Coeff.	$p > \|t\|$	Coeff.	$p > \|t\|$
Constant	0.6924 (0.0280)	0.000	0.6924 (0.0281)	0.000	0.1115 (0.0675)	0.099
Year (t)	−0.0579 (0.0462)	0.211	−0.0579 (0.0462)	0.211	−0.0996 (0.0388)	0.010
Treatment (T)	−0.0394 (0.0339)	0.246	−0.0394 (0.0339)	0.246	−0.0460 (0.0260)	0.077
Interaction	−0.0102 (0.0509)	0.841	0.1105 (0.0514)	0.032	0.0987 (0.0391)	0.012
Immig (interact)			−0.8959 (0.0567)	0.000	−0.6749 (0.0532)	0.000
Parental Occupation					0.0061 (0.0017)	0.000
Parental Education					0.0192 (0.0078)	0.014
School Type					0.0789 (0.0204)	0.000
School Resources					0.0138 (0.0087)	0.111
Village					0.0224 (0.0193)	0.246
Small Town					0.0134 (0.0188)	0.475
Large City					0.1115 (0.0675)	0.099

Source: Author's calculations using data from PISA (OECD 2004, 2010).
Note: SEs are presented in parentheses.

[13]See Heckman, Stixrud, and Urzua (2006), Lleras (2008) or Levin (2012) for a discussion on noncognitive education measures and the necessity of taking into account that dimension when educational achievement or labour market outcomes are being analysed.

[14]Simulations are based on the estimations from Models 2 (Equation 3) and 3 (Equation 4) contained in Tables 6 and 7. It makes no sense to run a simulation based on Model 1 because this model does not include the percentage of immigrants.

Table 8. Simulation of results for different percentages of immigrant students.

	All Students		Native Students	
% Immig	MODEL 2	MODEL 3	MODEL 2	MODEL 3
1	−0.01	−0.01	0.10	0.09
5	−0.03	−0.02	0.07	0.06
10	−0.06	−0.03	0.02	0.03
15	−0.08	−0.05	−0.02	−0.00
20	−0.11	−0.06	−0.07	−0.04
25	−0.14	−0.08	−0.11	−0.07
30	−0.17	−0.10	−0.16	−0.10
35	−0.19	−0.11	−0.20	−0.14
40	−0.22	−0.13	−0.25	−0.17
45	−0.25	−0.15	−0.29	−0.21
50	−0.28	−0.16	−0.34	−0.24

Source: Author's calculations.

percentage of enrolled immigrants has negative effects on the percentage of nonrepeaters for all students, although these effects are significant when the proportion of immigrants in the classroom is above 10%. For example, schools with a 10% concentration of immigrant students have around three immigrant pupils per classroom (for a 30-student classroom), which results in a decrease of from one to two nonrepeater pupils. In the case of native students, however, concentrations of immigrant students of under 15% have neither negative nor positive effects. Teachers appear to substitute potential native repeaters by these immigrant students when there are not many immigrant students in the class (fewer than four to five students) and the percentage of nonrepeating native students decreases.

However, when immigrant concentrations climb to over 15% (more than five immigrants per class), we start to detect a significant negative impact on natives' results compared with natives in the control group. According to the summary statistics presented in Table 3, this negative effect of immigrant concentration will impact on 34% of schools (those with immigrant concentration above 15% in 2009). In this case, the presence of six immigrant students per classroom (equivalent to an immigrant concentration of around 20%) leads to a reduction of from two to three individuals in the rate of nonrepeating native students. This finding, which is similar to previous findings reported in the literature (Calero and Waisgrais 2009), provides empirical evidence demonstrating that there is a clear negative peer effect related to a high concentration of immigrant students in some schools.

Placebo test

As mentioned in the methodology section, one assumption of the DiD method is that the trends of the treatment and control groups would be equal in the absence of the treatment, i.e. both groups are similar in all variables but the treatment. Because we cannot prove this assumption, we perform a *placebo* tests in order to check whether the identified effects are due to such treatment and endorse the correct selection of the control and treatment groups (Gertler et al. 2011).

In our research, we apply the *placebo* test using a fake-dependent variable – *average percentage of girls at school* –, knowing that it should not be affected by the increase of immigrant students in classrooms, but at the same time it seems to be correlated to grade retention, as girls are less likely to repeat a grade than boys (Corman 2003). Table 9 summarizes the results which corroborate our hypothesis: the DiD estimator (coefficient associated with the interaction term) and the DiD *dose* estimator (coefficient associated with immigrant concentration) are not statistically significant in any of the models.

V. Conclusions

Since the late 1990s, there has been a constantly growing inflow of immigrants, leading to a remarkable increase in the foreign population in Spain. This has affected the percentage of immigrant students who have joined the Spanish education system and account for around 9.5% of the school population for the year 2011 up from 1.5% in 2000. At the same time, Spain is feeling the effect of other relevant issues like consistently very high grade retention rates of around 30%.

Given this background, the aim of this article is to estimate the impact of the exogenous increase of immigrant students from 2003 to 2009 using a DiD approach, which would reveal whether immigrant concentration had a significant effect on the percentage of nonrepeater students. We use the data provided by consecutive OECD PISA reports.

In our identification strategy, schools with foreign students enrolled constitute our *treatment* group, whereas schools composed of only native students define our *control* group. On top of the traditional mean effect estimations, however, we analyse the impact of the concentration of immigrants in

Table 9. Placebo test: difference in differences models using *percentage of girls at school* as a fake dependent variable.

Dependent Variable	Model 1		Model 2		Model 3	
Percentage of Girls	Coeff.	$p > \|t\|$	Coeff.	$p > \|t\|$	Coeff.	$p > \|t\|$
Constant	48.3916 (1.5588)	0.000	48.3916 (1.5596)	0.000	54.4187 (4.7314)	0.000
Year (*t*)	1.2782 (1.6757)	0.446	1.2782 (1.6765)	0.446	1.9073 (1.6171)	0.238
Treatment (*T*)	0.6165 (1.9767)	0.755	0.6165 (1.9776)	0.755	0.3852 (1.8682)	0.837
Interaction	−0.1767 (2.1176)	0.934	−0.6931 (2.1648)	0.749	−0.4445 (2.0038)	0.825
Immig (interact)			3.8517 (3.1009)	0.214	−0.5403 (4.2394)	0.899
Parental Occupation					0.0483 (0.1259)	0.701
Parental Education					−0.5341 (0.3351)	0.111
School Type					−3.1862 (1.0709)	0.003
School Resources					0.6730 (0.4642)	0.147
Village					−1.3859 (1.5927)	0.138
Small Town					−1.6750 (1.2487)	0.180
Large City					1.5648 (4.7314)	0.539

Source: Author's calculations using data from PISA (OECD 2004, 2010).
Note: SEs are presented in parentheses.

classrooms in this article. For this reason, we refer to a *dose treatment* (Abadie and Dermisi 2008), where the dose is the percentage of immigrant students, and hence, the DiD estimator is the sum of the terms related to interaction and the percentage of immigrants ($\text{DD} = \beta + \delta\text{Immig}$).

Since we are interested in evaluating the effect of the immigration phenomenon on students and native students, in particular, we have two dependent variables: *percentage of students who are in their correct grade* and *percentage of native students who are in their correct grade*. For each dependent variable, we estimate three models: the basic DiD model (Model 1), an equivalent model introducing the treatment dose (Model 2) and an extension of the previous models that includes a set of control covariates (Model 3). Moreover, we develop a *placebo* test to check the validity and the robustness of the approach.

Analysing the effect on all students, we find that the interaction coefficient (β) (DiD basic impact estimator) appears not to be statistically significant; however, the term associated with the dose of immigrants (δ) (percentage of immigrant students) has a negative and statistically significant relationship with the percentage of students who are in their correct grade. The impact on native students is different, as the interaction coefficient (β) in the DiD dose estimator is statistically significant and positive, but this small advantage is offset and finally cancelled out by the *dose* term (δ) when the concentration of immigrants is above 15%.

In conclusion, immigrant students joining the Spanish education system does not, on average, decrease school promotion rates with respect to 2003. This situation is even beneficial to native students because foreign students are more greatly affected by grade retention. Taking into account the *dose* (percentage of immigrants enrolled per school), however, we find that the concentration of immigrant students has a negative impact on promotion rates. In other words, the average percentage of repeaters, and, in particular, the average percentage of native repeaters, has increased in 2009 with respect to 2003 as a consequence of higher immigrant concentrations in some schools. However, native students are only affected by higher concentrations of immigrant students (above 15%).

The key question is why the addition of immigrant students had such an impact on the education system. A potential reason for this result is that immigrant students have a language deficit and lower educational level when they join the Spanish education system. Therefore, when the number of immigrant students per classroom grows, the average educational level of the students in these classrooms drops and more students fail to reach the educational level for promotion. Some possible educational strategies to manage this situation would be to regulate the maximum percentage of immigrants per school in order to avoid high concentrations. Nevertheless, once high concentrations of immigrant pupils is a fact in some schools, policy-makers should contemplate specific strategies in order to avoid the negative effects of large concentration of immigrant students and at the same time, fostering the improvement of immigrant students' educational attainment. Policies such as the provision of more resources for specific language and skills training in those schools with high concentration of immigrants enrolled could solve problems of adaptation to the new education system. Those resources could be employed for hiring specialized teachers who focus on immigrant students and their progress or for reducing size of classrooms so the concentration of immigrant pupils would be lower. Moreover, it is widely known in the literature of economics of education that both family background and home environment are key variables to children' learning process; hence, policies aimed to improve the integration of those immigrant families would be extremely useful. Strategies driven to reduce their labour market insertion problems or special instructions for acquisition of the new language would impact positively on immigrant socioeconomic status and, as a consequence, on their offspring's educational performance.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Extremadura Government [IB13106].

References

Abadie, A. 2005. "Semiparametric Difference-In-Differences Estimators." *Review of Economic Studies* 72: 1–19. doi:10.1111/roes.2005.72.issue-1.

Abadie, A., and S. Dermisi. 2008. "Is Terrorism Eroding Agglomeration Economies in Central Business Districts? Lessons from the Office Real Estate Market in Downtown Chicago." *Journal of Urban Economics* 64: 451–463. doi:10.1016/j.jue.2008.04.002.

Ammermueller, A. 2007. "Poor Background or Low Returns? Why Immigrant Students in Germany Perform so Poorly in the Programme for International Student Assessment." *Education Economics* 15: 215–230. doi:10.1080/09645290701263161.

Benito, A. 2007. "La LOE ante el fracaso, la repetición y el abandono escolar." *Revista Iberoamericana De Educación* 43: 1–11.

Brunello, G., and L. Rocco. 2013. "The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores." *Economics of Education Review* 32: 234–246. doi:10.1016/j.econedurev.2012.10.006.

Calero, J., A. Choi, and S. Waisgrais. 2009. "Determinantes del rendimiento educativo del alumnado de origen nacional e inmigrante." *Cuadernos Económicos Del ICE* 78: 281–311.

Calero, J., and S. Waisgrais. 2009. "Rendimientos educativos de los alumnos inmigrantes: identificación de la incidencia de la condición de inmigrante y de los peer-effects." Proceedings of the XVI Meeting of Public Economic, Granada, Spain.

Cebrián, J. A., M. I. Bodega, M. A. Martín-Lou, and F. Guajardo. 2010. "La crisis económica internacional y sus repercusiones en España y en su población inmigrante." *Estudios Geográficos* 71: 67–101. doi:10.3989/egeogr.2010.i268.

Chaudhury, N., and D. Parajuli. 2010. "Conditional Cash Transfers and Female Schooling: The Impact of the Female School Stipend Programme on Public School Enrolments in Punjab, Pakistan." *Applied Economics* 42: 3565–3583. doi:10.1080/00036840802167376.

Contini, D. 2013. "Immigrant Background Peer Effects in Italian Schools." *Social Science Research* 42: 1122–1142. doi:10.1016/j.ssresearch.2013.02.003.

Corman, H. 2003. "The Effects of State Policies, Individual Characteristics, Family Characteristics, and Neighbourhood Characteristics on Grade Repetition in the United States." *Economics of Education Review* 22: 409–420. doi:10.1016/S0272-7757(02)00070-5.

Cristia, J., A. Czerwonko, and P. Garofalo. 2014. "Does Technology in Schools Affect Repetition, Dropout and Enrollment? Evidence from Peru." *Journal of Applied Economics* 17: 89–111. doi:10.1016/S1514-0326(14)60004-0.

Driessen, G. W. J. M. 2001. "Ethnicity, Forms of Capital and Educational Achievement." *International Review of Education* 47: 513–538. doi:10.1023/A:1013132009177.

Entorf, H., and N. Minoiu. 2005. "What A Difference Immigration Policy Makes: A Comparison of PISA Scores in Europe and Traditional Countries of Immigration." *German Economic Review* 6: 355–376. doi:10.1111/geer.2005.6.issue-3.

Felfe, C., N. Nollenberger, and N. Rodríguez-Planas. 2015. "Can't Buy Mommy's Love? Universal Childcare and Children's Long-Term Cognitive Development." *Journal*

of Population Economics 28: 393–422. doi:10.1007/s00148-014-0532-x.
Frick, J. R., and G. G. Wagner 2001. “Economic and Social Perspectives of Immigrant Children in Germany.” *Discussion Paper 301*, IZA.
Ganzeboom, H. B. G., P. M. De Graaf, and D. J. Treiman. 1992. “A Standard International Socio-Economic Index of Occupational Status.” *Social Science Research* 21: 1–56. doi:10.1016/0049-089X(92)90017-B.
Geay, C., S. McNally, and S. Telhaj. 2013. “Non-native Speakers of English in the Classroom: What Are the Effects on Pupil Performance?” *The Economic Journal* 123: F281–307. doi:10.1111/ecoj.12054.
Gertler, P. J., S. Martinez, P. Premand, L. B. Rawlings, and C. M. J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: World Bank.
Heckman, J. J., J. Stixrud, and S. Urzua 2006. “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behaviour.” *Working Paper 12006*, National Bureau of Economic Research.
Jimerson, S. R., G. E. Anderson, and A. D. Whipple. 2002. “Winning the Battle and Losing the War: Examining the Relation between Grade Retention and Dropping Out of High School.” *Psychology in the Schools* 39: 441–457. doi:10.1002/(ISSN)1520-6807.
Khandker, S. R., G. B. Koolwal, and H. A. Samad. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, DC: World Bank.
Levin, H. M. 2012. “More than Just Test Scores.” *Prospects* 42: 269–284. doi:10.1007/s11125-012-9240-z.
Lleras, C. 2008. “Do Skills and Behaviors in High School Matter? the Contribution of Noncognitive Factors in Explaining Differences in Educational Attainment and Earnings.” *Social Science Research* 37: 888–902. doi:10.1016/j.ssresearch.2008.03.004.
Meyer, B. D. 1995. “Natural and Quasi-Experiments in Economics.” *Journal of Business & Economic Statistics* 13: 151–161.
Murat, M., and P. Frederic. 2015. “Institutions, Culture and Background: The School Performance of Immigrant Students.” *Education Economics* 23: 612–630. doi:10.1080/09645292.2014.894497.
OECD. 1999. *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*. Paris: OECD.
OECD. 2004. *Learning for Tomorrow’s World. First Results from PISA 2003*. Paris: OECD.
OECD. 2005. *PISA 2003 Data Analysis Manual. SPSS Users*. Paris: OECD.
OECD. 2009a. *PISA 2006 Technical Report*. Paris: OECD.
OECD. 2009b. *PISA Data Analysis Manual. SPSS*. 2nd ed. Paris: OECD.
OECD. 2010. *PISA 2009 Results: What Students Know and Can Do*. Paris: OECD.
OECD. 2011. *How are School Systems Adapting to Increasing Numbers of Immigrant Students? PISA IN FOCUS*. Paris: OECD.
Ohinata, A., and J. C. Van Ours. 2013. “How Immigrant Children Affect the Academic Achievement of Native Dutch Children.” *The Economic Journal* 123: F308–331. doi:10.1111/ecoj.12052.
Ravallion, M. 2008. “Evaluating Anti-Poverty Programs.” In *Handbook of Development Economics*, edited by T. P. Schultz and J. Strauss. Amsterdam: North-Holland.
Roderick, M. 1994. “Grade Retention and School Dropout: Investigating the Association.” *American Educational Research Journal* 31: 729–759. doi:10.3102/00028312031004729.
Rojas, J., and P. Sánchez. 2011. “Evolución reciente y proyecciones de la población en España.” *Boletín Económico del Banco de España* 6: 61–68.
Rutkowski, L., E. Gonzalez, M. Joncas, and M. Von Davier. 2010. “International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting.” *Educational Researcher* 39: 142–151. doi:10.3102/0013189X10363170.
Salinas, J., and D. Santín. 2012. “Selección escolar y efectos de la inmigración sobre los resultados académicos españoles en PISA 2006.” *Revista de Educación* 358: 382–405.
Sánchez, P. 2013. “Evolución reciente de la población en España y proyecciones a corto y largo plazo.” *Boletín Económico del Banco de España* 1: 71–80.
Schnepf, S. V. 2008. “Inequality of Learning Amongst Immigrant Children in Industrialized Countries.” *Discussion Paper 3337*, IZA.
Turner, R. 2006. “El Programa Internacional para la Evaluación de los Alumnos (PISA). Una perspectiva general.” *Revista de Educación*, 2006: 45–74.
Zinovyeva, N., F. Felgueroso, and P. Vazquez. 2014. “Immigration and Student Achievement in Spain: Evidence from PISA.” *SERIEs* 5: 25–60. doi:10.1007/s13209-013-0101-7.