Working papers series

WP ECON 16.09

No student left behind? Evidence from the Programme for School Guidance in Spain

J.I García-Pérez (Universidad Pablo de Olavide)

Marisa Hidalgo-Hidalgo (Universidad Pablo de Olavide)

Keywords: Remedial education, PAE, programme evaluation, PISA, selection bias.

JEL Classification: H52, I23, I28, J24

Department of Economics

No student left behind? Evidence from the Programme for School Guidance in Spain*

J. Ignacio García-Pérez
Universidad Pablo de Olavide & FEDEA

Marisa Hidalgo-Hidalgo
Universidad Pablo de Olavide

September 13, 2016

Abstract

This paper evaluates the effects of a remedial education programme implemented in Spain between 2005 and 2012 that offered after-school classes for underperforming students from poor socioeconomic backgrounds. We use two different estimation strategies, re-weighting estimators and propensity score matching, and address the existence of selection bias. We find that this programme had a substantial positive effect on children's academic achievement: the probability of falling behind the general progress of the group declined by approximately 5% and mean reading scores increased by approximately 10% of one standard deviation. We also find that a larger exposure to the programme improves students' scores: whereas students in schools that participated in the programme for at most two years do not experience any significant positive effect, those in schools that participated for at least three years did. The programme significantly reduced the probability of belonging to the bottom part of the distribution (by approximately 7.5%) and improved mean scores (by approximately 18% of one standard deviation). Finally, we find that the impact of the programme is much stronger for students in rural schools than for students in urban schools.

Keywords: Remedial education, PAE, programme evaluation, PISA, selection bias
JEL Classification: H52,I23,I28,J24

*We would like to thank Laura Hospido and Ernesto Villanueva for their very helpful comments and suggestions. We also thank seminar participants at Alicante, Salamanca and the 40 Simposio de Análisis Económico. We gratefully acknowledge Ismael Sanz and Francisco Javier García Crespo (INEE) for their help with the data used in this article. We are also grateful for the assistance with the dataset offered by Angelica Martínez Zarzuelo (INEE). Financial support from the Spanish Ministry of Innovation and Science [ECO2013-43526-R and ECO2014-57413] and Junta de Andalucia [SEJ-5980; SEJ-426, P09-SEJ6882] is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

Growing evidence shows that inequality has increased in many developed countries in recent decades.[1] Recent OECD data (OECD, 2013) indicate that the global economic crisis reduced incomes and that this reduction is not shared evenly across the income distribution, as there are larger reductions in the bottom, thus suggesting further increases in inequality and poverty. In addition to the global crisis, this evidence might also reflect the fact that both low-skilled workers and low-achieving students are being left behind by rapid technological change in a globalized world economy (see Freeman, 2008 or Kanbur, 2014). Indeed, poor-achieving students are more likely to be early school leavers, which has long-run negative effects, increasing the risk of social exclusion and poverty.[2] This recent evidence has arguably made improving the education and skills of the workforce a priority, impelling policy makers to address poverty and exclusion and promote growth. Indeed, one of the EU's education targets for 2020 is to reduce the rates of young people leaving early education and training. In addition, the European Union's 2013 Social Investment Package focusses on policies designed to strengthen people's skills and capacities, including education and childcare, as well as active labour market policies (see European Commission 2013a and 2013b). These developments leave us with the following question: how do we make education a success for disadvantaged students in developed countries?

Remedial education programmes are designed to help poor-performing students to satisfy minimum academic standards. This is usually achieved by means of a targeted increase in instruction time combined with after-school individualized instruction in small study groups. Therefore, these types of interventions are currently subject to increasing interest. While remedial education is quite widespread in the U.S., there is less of a tradition in Europe.[3] Moreover, the evidence on the effectiveness of such programmes is scarce. Providing such evidence is precisely the goal of this paper. Namely, our objective is to evaluate the effects of a multiyear programme implemented in Spain between 2005 and 2012 that offered remedial education for underperforming students from poor socioeconomic backgrounds. This remedial programme is the Programme for School Guidance (PAE, which is the Spanish acronym for *Programa de Acompañamiento Escolar*). In particular, we attempt to address the following two questions: does the programme reduce the number of students left behind the general progress of the group? Does the programme improve students' mean scores? We assess whether the intervention succeeded in achieving these two goals while it was being implemented (we refer to this as the PAE-Immediacy treatment). In addition, we analyse whether the programme was more effective in achieving both objectives the longer a school participated in it. To do so, we use external evaluations of the schools: the PISA 2012

[1] See, among others, Atkinson, 2010 for the EU and Atkinson et al, 2011 for the US.

[2] See Brunello and De Paola (2014) and references therein for a review of the private and social cost of early school leaving in Europe.

[3] See the European Commission (2013) for a review of remedial programmes in Europe. Among several examples of remedial interventions at school in the U.S., see, for example, the Bell After-School Instructional Curriculum (BASICs) or that promoted by the 21st Century Community Learning Centers (U.S. Department of Education, 2003).

exams.[4]

Our main results suggest that the PAE had a substantial positive effect on students' academic achievement. It reduced the probability of falling behind into the bottom part of the reading score distribution by approximately 5% (nearly 10% of one standard deviation). The estimated effect on mean reading scores is above 12 PISA points (more than 14% of one standard deviation). We also find that a larger exposure to the programme improved students' scores: whereas students in schools that participated in the programme for at most two years do not experience any significant positive effect, those in schools that participated in the programme for at least three years did. The PAE significantly reduced the probability of belonging to the bottom part of the distribution (by approximately 7.5%) and improved mean scores (by approximately 18% of one standard deviation). Furthermore, our evidence suggests that there is heterogeneity in the impact of the programme across school types, namely, urban versus rural. In particular, we find that the impact of the programme is much larger among students attending rural schools than students attending urban schools (according to several indicators, the impact in rural schools is more than twice that for urban schools).

Remedial programmes are often very difficult to evaluate due to sample selection. Students' individual and socioeconomic characteristics affect both their probability of being selected for the programme and its success, as the selection mechanism is not completely observable. Fortunately, the richness of our data, combined with access to schools' performance in 2009 (before a group of schools joined the programme) and in 2012 (after joining it) allows us to control for a variety of observable student characteristics and address unobservables that might affect the selection of schools for the PAE and their outcomes. Our first estimation strategy compares the PISA 2012 reading scores of those students that attended schools that participated in the PAE with the hypothetical outcome that these same students would have obtained had they not attended PAE schools. The counterfactual reading score is inferred using a control group composed of students in schools that did not join the PAE but participated in PISA 2012. To ensure that treatment and control groups are comparable on observables, students in the control group are re-weighted by assigning relatively more weight to those students whose individual, family and school characteristics are similar to those of the means of the treated group. As a second estimation strategy, we propose using propensity score matching to examine the impact of the PAE. In addition, we estimate the role of unobservable variables in the schools' decision to volunteer for the PAE. The availability of information on student performance in schools before joining the programme allows us to examine the existence of selection bias. This is one of the contributions of this paper.[5] We estimate the selection bias by combining, on the one hand, the information available in PISA 2009 exams with, on the other, the information regarding participation in the PAE one, two or three years later. We identify in the PISA 2009 sample those schools that volunteered

[4]PISA is the Programme for International Student Assessment. It measures students' skills in three areas: mathematics, reading and science.

[5]See also Hospido et al. (2015) who employ a similar approach to examine the impact of a financial education programme on students' scores.

for the PAE only after 2009. In this sample, any difference in reading performance among students in schools that volunteered for the PAE only after 2009 and those in schools that never participated in the PAE can be attributed solely to the existence of selection bias. We do not find any significant selection bias. A possible explanation is that, as the programme began during the 2005/06 academic year, by the 2009/10 academic year, and afterwards, the existence of the programme was quite widespread in the education community (the rate of programme participation exceeds 45% in some regions).

Our paper contributes to the relatively scarce literature on the evaluation of remedial education programmes for teenage students in developed countries.[6] Only a few works address the identification problem and obtain evidence regarding the effectiveness of these programmes in the short run. Jacob and Lefgren (2004) analyse the effect of summer schools on the performance of 9-12 year-old students in Chicago and find that the net effect of these programmes was to substantially increase academic achievement among third-graders but not sixth-graders. Lavy and Schlosser (2005) evaluate the short-term effects of the Bagrut 2001 programme, a remedial intervention very close in spirit to that evaluated in this study, which provided additional instruction to underperforming high school students in Israel. Their study shows that it was more cost effective than alternatives based on financial incentives for pupils and teachers. Holmlund and Silva (2014) study a remedial education programme targeting English secondary school pupils at risk of school exclusion that, instead of targeting standard cognitive skills (as does the PAE and other programmes mentioned above), targeted students' non-cognitive skills, finding little evidence that the programme significantly helped treated youths to improve their age-16 test outcomes. A recent contribution is Battaglia and Lebedinski (2015), who analyse the impact of the Roma Teaching Assistant Programme in Serbia. However, their work differs somewhat from our study, as it is focused on a stigmatized ethnic group, Roma pupils. Thus, one of the contributions of this paper is that it is among the first to analyse the impact of a remedial education programme on students' academic achievement within the European context, which is crucial considering the current debate over the increasing inequality and poverty in Europe.[7] Therefore, our insights might be highly relevant from a policy perspective.

The present paper is organized as follows. Section 2 summarizes the PAE and presents the data and descriptive statistics used in the paper. Section 3 describes the methodology. Section 4 reports the results. In Section 5, we examine the existence of selection bias. Section 6 provides a robustness check of the main results. Finally, Section 7 concludes.

[6] Evidence on the impact of remedial and analogous programme in developing countries is more common. See, for example, Banerjee et al. (2007), who evaluate the Balsakhi Programme in India or, more recently, Kremer et al (2013) for a review of the existing evidence on programme impact in developing countries.

[7] A number of recent papers have focused on remedial programmes in tertiary education in Europe and the U.S. For example, De Paola and Scoppa (2014) and De Paola and Scoppa (2015) analyse the impact of remedial courses on the achievement of college students in Italy. Bettinger and Long (2009) and Calcagno and Long (2008) study the causal effect of remediation on the outcomes of college students in Ohio and Florida, respectively.

2 The PAE

The Spanish education system is organized into three levels: primary (grades 1-6), secondary (grades 7-10) and pre-college (grades 11-12). The first two levels are compulsory (a student can choose to leave school at age 16). Most schools provide either primary or secondary and pre-college education.[8] All students born in the same calendar year must enter school in the same academic year, with 10th grade being the reference grade for 15-year-old students (who are the students in our sample).

The PAE is a programme targeting public primary and secondary schools. It was implemented during the period 2005-2012. The PAE is an example of a set policies implemented in Spain to improve poor educational outcomes: the early drop-out rate was over 30% in 2004 and 2008 (see Spanish Ministry of Education, 2016). The PAE provides support to public schools with a significant number of students from disadvantaged backgrounds.[9] The aim of this intervention was to enhance the learning abilities and academic returns of underperforming students with poor socioeconomic backgrounds. This was pursued by stimulating reading habits, providing students with study organization techniques, and improving their social abilities. It consisted of providing support (at least 4 hours per week) during after-school hours to those students with special needs and learning difficulties. This support was provided after-school by instructors or teachers from the students' own schools who worked with these students in small groups (5-10 students). Students were selected by both their tutor and the rest of the teachers and could be in any grade within secondary school. They were chosen based on their poor academic results, general motivation and prospects, although there was no single quantifiable and explicit selection rule. During the remedial classes, the students engaged in guided reading and worked on the subjects that presented particular difficulties for them. Instructors offered clarification, provided additional material, assisted students with work organization techniques, etc.[10] Although the PAE was implemented in both primary and secondary schools, we focus our analysis on secondary schools. The reason is that we use PISA 2012 exam results as the means of evaluating the PAE, and this exam is taken by 15-year-old students, with 10th grade being the reference grade for them). Finally, as the programme was implemented only in public schools, we exclude from the PISA database both private and private but publicly financed schools.

Figure 1 displays the percentage of public secondary schools in which the PAE was implemented in each region during the full period that the programme was implemented, that is, from the 2005/06 until the 2011/12 academic year. We distinguish five sub-periods, displayed in the five panels in Figure 1. In panels 2 to 5, we have the four academic years that the student attended the same secondary school where she took the PISA exams, that is, 2008/09, 2009/10, 2010/11 and 2011/12, meaning grades 7 to 10. In panel 1, we include the

[8] Only a very small sample of schools (most of them private) provide the three levels. See Spanish Ministry of Education (2016).

[9] Schools volunteered for the programme and committed themselves to improving their students' outcomes by providing after-school instruction to those students with special needs.

[10] For additional details on the PAE, see (only in Spanish) http://www.mecd.gob.es/educacion-mecd/areas-educacion/comunidades-autonomas/programas-cooperacion/plan-proa/acompanamiento-escolar-secundaria.html

preceding years, 2005 through 2008, that is, the period in which the PAE might also have been implemented at that school but, as the students in our sample are 15-years-old ones, they did not benefit from it since they were attending primary courses at a different school during that period.[11]

Here Figure 1: Schools with the PAE

The figure indicates that the PAE was progressively introduced throughout this period. The percentage of schools participating in the PAE was very low during the first three academic years (below 1% in most regions). However, during the period analysed in this paper from 2008 until 2012, there was a gradual implementation of the programme in most regions (the proportion of schools with the PAE was above 40% in several regions).[12]

Next, we analyse how the PAE was introduced in the schools in our sample during the 2005-2012 period. We consider the same five sub-periods mentioned above: 2005/08, 2008/09, 2010/11, 2010/11 and 2011/12. Thus, depending on whether the school implemented the PAE in each of these sub-periods, we may have 28 different types of schools. Table 1 below reports the number of schools of each type. It also shows some descriptive statistics for the schools (their mean reading PISA scores and an index of economic, social and cultural status, ESCS). For example, the first thirteen rows show the number of schools where the PAE was implemented at least during the last academic year we consider, 2011/12, regardless of whether it was implemented before.

Here Table 1: PAE implementation

Several comments can be made. Most schools in our sample, more than 60%, did not implement the PAE. Among those that did, the majority implemented the programme throughout the period considered. For example, the first row indicates that more than 10% of the schools in our sample participated in the PAE during every academic year (from 2005 until 2012). Moreover, once a school joins the programme, it is very likely to continue participating in it. For example, only 7 schools in the sample participated in the PAE from the very beginning (2005/06 academic year) but dropped out during the last academic year. Finally, schools where the PAE was implemented for a longer period do not seem to differ from other schools in terms of mean reading achievement or the socioeconomic index. For example, both the mean reading score and ESCS of schools where the PAE was implemented throughout the full period are not statistically different from the corresponding means for the full sample. This suggests that schools joined the programme in no particular order. In Section 5 below,

[11] Some secondary schools also teach primary education levels (see Footnote 8). Moreover, these schools (and, thus, the students there) might have participated in the PAE at primary level during this period (from 2005 until 2008). Nevertheless, as we cannot identify these schools in our database, we will simply consider as treated students those attending secondary schools that joined that programme at that level.

[12] As explained before, schools volunteered for the programme. They received funding and had to manage programme implementation. The criteria to distribute funds for the programme among regions included the number of public schools, the number of students attending public schools and the number of early school leavers or dropouts. Apparently, the guidelines to distribute funds among schools within regions resemble the previous iterations: as we will note below, both a school's size and its proportion of dropouts increase the probability of joining the programme.

we examine this hypothesis.

As noted above, we use external evaluations of schools, specifically, PISA 2012 scores for the regions with enlarged samples.[13] There are at most 35 students per school participating in PISA. These students are selected based on a two-stage sample design developed by the PISA programme organizers. This selection ensured representation of the full target population of 15-year-old students in the participating countries.[14] Table 2 below shows, first, the number of secondary schools that participated PISA in 2012 per region (column 1). It also shows the number of secondary schools where the PAE was implemented in a particular academic year, regardless of whether it was also implemented in other academic years (see, for example, column 2 for the figures corresponding to the 2005/06 academic year). Finally, it shows the number of schools where the PAE was implemented and that also took PISA 2012 (see column 3, for example, for the figures corresponding again to the 2005/06 academic year). As can be observed, more than 10% of the schools where the PAE was implemented during 2011/12 were also evaluated in PISA 2012 (see columns 14 and 15).

Here Table 2: Schools with the PAE in PISA 2012

The PISA 2012 database provides individual-level information on demographics (e.g., gender, immigration status, month of birth), socioeconomic background (parental education and occupation), school-level variables and achievement test scores in three disciplines: science, maths and reading. We focus on test scores in reading, as the PAE focussed primarily on improving learning abilities by stimulating reading habits, as noted above. Nevertheless, we also assess the impact of the PAE on science and maths scores. In addition, as the main goal of the PAE was to improve poor educational outcomes among students from disadvantaged backgrounds, we concentrate our analysis on the performance of that specific group of students. In particular, we define as our main outcome variable the probability of falling behind the general progress of the group or being a low achiever. In doing so, we use the score in the first quartile for reading to define the group of "lowest achievers". Additionally, we also consider as an outcome variable the student's reading score.

Finally, we do not consider in the analysis schools that joined other remedial programmes.[15]

[13] The regions with a representative (enlarged) sample are Andalusia, Aragon, Asturias, Balearic Islands, Canary Islands, Cantabria, Castile Leon, Catalonia, Extremadura, Galicia, La Rioja, Madrid, Murcia, Navarre, and Basque Country. The 2012 edition of PISA focused on science. Following the OECD's recommended methodology, we use the 5 plausible values in the PISA Technical Report to calculate each student's educational outcome.

[14] Only in a few cases, and with proper justification, PISA national project managers can exclude certain schools (e.g., in a remote geographical region) or students (e.g., special needs students). Nevertheless, the guidelines explicitly state that students must not to be excluded solely because of poor academic performance or normal discipline problems. See the PISA 2012 Technical Report for further details.

[15] The PAE is part of a larger remedial programme: PROA (which is the Spanish acronym for Plan de Refuerzo, Orientación y Apoyo, literally, Plan for Reinforcement, Guidance and Support). In addition to the PAE, some schools also participated in another PROA-related programme: PAR (which is the Spanish acronym for Programa de Apoyo y Refuerzo, literally, Programme for Reinforcement and Support). It consists of providing additional resources to schools. We focus on the PAE because both the target population and the intervention are more clearly defined: the target population of the PAE is students with poor academic results, whereas in PAR, it is not only students but also their parents and the school in general. Second, the

Our final sample consists of 11,747 individuals from 417 schools. We refer to this as our evaluation sample. Table 3 reports the main descriptive statistics of a set of individual, socioeconomic and school-level variables for the evaluation sample (in column 1) and for all public schools in the PISA sample (column 2), that is, schools that joined other remedial programmes, in particular PAR (see Footnote 14 for a description of PAR and Appendix 1 for a detailed definition of the variables in the paper):

Here Table 3: Summary Statistics

As Table 3 indicates, the mean reading score for students in the evaluation sample is higher than that for all public schools. The proportion of immigrants and repeaters is lower in the evaluation sample.[16] However, there is no difference in the proportion of girls or in the proportion of students who attended pre-primary schools for more than one year. Regarding socioeconomic characteristics, our evaluation sample have a slightly smaller proportion of students from disadvantaged families: both the proportion of students with an educated father or mother and the index of educational items in the home are higher than in the full sample of public schools. Finally, we observe that the socioeconomic composition of the schools in the evaluation sample is quite similar to the full sample of public schools: the proportion of students with educated parents and the mean socioeconomic index at the school level are very similar in both samples. The proportion of dropouts in the schools in the evaluation sample is lower than in the full sample of public schools. In addition, students in the evaluation sample are in smaller schools and more likely to be in rural areas.

Next, we comment on the design of the programme evaluation. The programme was implemented for several years, and thus we can consider many different treatment definitions (see Table 1). Most students in the sample attended the same school for at least the most recent four academic years prior to taking the PISA exam in 2012, that is, 2008/09, 2009/10, 2010/11 and 2011/12. Therefore, they could be treated in any of these academic years. We focus here on the primary or initial effect of that programme. Thus, we consider as *treated* students those at schools that participated in the PAE during the same academic year in which PISA exams were taken, namely, 2011/12, regardless of whether the school joined the programme before (that is, in any academic year between 2005/06 and 2010/11). We consider as *controls* students in schools where the PAE was not implemented at all (that is, in any academic year between 2005/06 and 2011/12). We drop from the analysis students in schools where the PAE was implemented during any academic year between 2005/06 and 2010/11 but not thereafter, i.e., during 2011/12. We refer to this treatment as *PAE-Immediacy*.

intervention in the PAE is similar across schools (providing students with additional classes), whereas under PAR, this was not always the case (improving school infrastructure, follow families more closely, etc.). To provide cleaner results, we drop from the analysis those schools that, in addition to the PAE, also joined the PAR programme. The total budget for the PROA Programme in 2005 was 8.5 million euros, whereas it was in excess of 400 million euros in 2012, the last year it was implemented. See the Spanish Ministry of Education Website at http://www.mecd.gob.es/educacion-mecd/areas-educacion/comunidades-autonomas/programas-cooperacion/plan-proa.html

[16] See García-Pérez et al. (2014) for a detailed analysis of the impact of being a repeater on student achievement.

In addition, we also assess whether the impact of the programme is stronger the more years it was implemented. To do so, we define two different treatments and compare their results. We first consider as treated students those at schools where the PAE was implemented for only one or two of the last four academic years. We refer to this treatment as *PAE-Intensity 1-2 years*. Second, we consider as treated students those at schools where the PAE was implemented for three or four of the last four academic years. We refer to this treatment as *PAE-Intensity 3-4 years*. Similar to *PAE-Immediacy*, as controls in the previous two treatments, we employ students in schools where the PAE was not implemented at all (that is, in any academic year between 2005/06 and 2011/12). We refer to the comparison between the results of these two treatments as *PAE-Intensity*. See Table 4 below for a summary of the several treatment definitions.

Here Table 4: Treatment definitions

Table 5 reports the number of treated and control schools in the sample according to each of the treatments defined above. The number of control schools (and students) is the same in the three treatments previously defined. The high survival rate of the PAE might explain why the number of treated schools in the 3-4 year treatment is larger than that in the 1-2 year treatment.

Here Table 5: Treated and control: school and students.

Table 6 reports descriptive statistics for the treated and control groups and balancing tests corresponding to our main treatment definition, *PAE-Immediacy*: treated students (column 1), control (column 2) and the difference between the two (column 3).

Here Table 6: Summary Statistics: Treated and controls

Mean reading test scores are lower among students in treated schools than among students in control schools. In Table 6, we also report the percentage of treated and non-treated students whose reading scores are below the first quartile (P25) in the corresponding score distribution (Reading25). The percentage of low-performing students is larger in the treatment group. There are not significant differences with respect to gender composition between the two groups. However, students in PAE schools differ from those in schools that did not join the programme: control students are less likely to be immigrants and are 4 points less likely to have repeated a grade. In addition, the proportion of educated parents (mother and/or father), the index of educational materials and the mean socioeconomic index are lower among treated students, suggesting that treated schools have a higher proportion of students from disadvantaged backgrounds. Finally, treated students came from larger sized schools and exhibited a larger proportion of dropouts. Conversely, students in the control sample are from schools with a higher student-teacher ratio and with principals that more frequently work to enhance the school's reputation in the community. In the analysis below, we comment on weighted treated and control students in columns (4) and (5) of Table 6.

3 Empirical Strategy

We study the effects of the PAE on the students' probability of falling behind the general progress of the group (having a score in the first quartile in the reading score distribution) and students' reading score, considering the students as the unit of analysis. By selecting the student as the unit of observation, we are aware that, to the extent that we cannot observe whether a particular student actually received the treatment, we can only consider them *potentially* treated, and thus, the effect we study in this case is the *potential* effect of the PAE. Nevertheless, we address this point below and attempt to provide a cleaner estimate of the true effect of the PAE by decomposing our evaluation sample.[17]

In the evaluation literature, data often come from non-randomized studies. The main assumption in this literature is that individuals' participation in the policy intervention to be studied can be considered a random event or, at least, independent of treated and control individuals' characteristics (see Myoung-Jae Lee, 2005). However, selection into the treatment is not independent of treated and control individuals' characteristics. In their seminal work, Rosenbaum and Rubin (1983) proposed propensity score matching as a method to reduce the bias in the estimation of treatment effects when using such datasets. This method consists of performing a matching between individuals (students) in the treatment group and individuals in the control group who are as similar as possible with respect to observables (individual, socioeconomic and school variables).[18] This implies dividing the sample into cells containing very similar individuals. However, if the vector of observable characteristics is too large, it is possible that we may lack sufficient observations from treated and control individuals with exactly the same values for every control variable. That is, there is not a positive number of observations within each cell. Propensity score matching is a way to "correct" this problem. The propensity score is defined by Rosenbaum and Rubin (1984) as the probability of being treated considering those variables included in the set of regressors.[19] This method proposes to summarize the pre-treatment characteristics of each subject into a single-index variable (the propensity score) that makes the matching feasible. This index is built based on the estimation of the probability of being treated, $p(X_i)$, where X_i denote the vector of pre-treatment characteristics. If D_i denote a binary variable that indicates exposure

[17] In addition, we study the impact of the programme while considering the school to be the treatment unit as a robustness check and obtain results that are qualitatively unchanged and very similar in size (see Section 6).

[18] Heckman et al. (1998) proposed three factors that contribute to reduce selection bias in a evaluation study. First, we need pre-treatment variables. Second, all of the information should come from the same data source. Third, both populations (treated and non-treated) must be in the same geographical area. Our study satisfies the first two conditions due to the specific characteristics of our dataset: academic scores come from PISA, and data on school participation in the PAE come from both regional and Ministry of Education registers. We believe that our study satisfies the third condition due to the low mobility within regions in Spain.

[19] Any standard probability model can be used to estimate the propensity score. The dependent variable is a binary variable equal to 1 if the individual has been treated or 0 if he has not been treated. Either a logistic distribution (logit model) or a normal distribution (probit model) may be used.

to the treatment:

$$D_i = \begin{cases} 1 & \text{if treated} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

and, as mentioned above, the definition of the treated group depends on the specific treatment considered: *PAE-Immediacy* or *PAE-Intensity* (1-2 years or 3-4 years),[20] the propensity score is defined as the conditional probability of PAE "participation" given pre-treatment characteristics, X:

$$p(X_i) \equiv Pr(D_i = 1|X) = E(D|X) \tag{2}$$

3.1 Re-weighting estimates

Now, let Y_i^1 denote the potential outcome (PISA reading score or probability of falling into the first quartile of the reading distribution) that student i would have obtained had she received the PAE treatment and Y_i^0 had she not received the PAE treatment. We denote by Y_i the PISA outcome, and thus, $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$. Therefore, the average effect we are interested in estimating when evaluating the PAE is

$$\tau = E(Y_i^1/D = 1, X) - E(Y_i^0/D = 1, X), \tag{3}$$

where the second term is the counterfactual outcome in the absence of the treatment and, thus, is unobservable and must be estimated. This is achieved using the outcomes of control students (that is, those in schools where the PAE was not implemented at all). This requires that the characteristics of the control and treatment group be as similar as possible. However, as previously mentioned, treated and control students differ with respect not only to their demographic characteristics, but they also differ in socioeconomic background and attend different schools (see Table 6). To solve this problem, we use the rich information on demographic, parental and school characteristics in the PISA 2012 database to *re-weight* the sample of controls such that they can provide a counterfactual to the PISA scores of the treated students. Formally, under the standard assumptions of conditional independence or unconfoundedness:

$$(Y_i^1, Y_i^0) \perp D_i \mid X_i \tag{4}$$

that is, within each cell defined by X_i, treatment is random, or similarly, the selection into treatment depends only on the observables X_i and common support:

$$p(X_i) \in (0, 1) \tag{5}$$

we have that:

$$E(Y_i^0/D = 1, X) \equiv E(\omega(x_i)Y_i/D = 0, X) \tag{6}$$

where $\omega(x_i) = \frac{1-\pi}{\pi} \times \frac{p(X_i)}{1-p(X_i)}$ and $\pi = \Pr(D_i = 1)$.

This expression indicates that we can identify the mean impact on treated individuals were

[20] Here we consider as controls students in schools where the PAE was not implemented at all, i.e., between 2005 and 2012 (see Table 4 above for details).

they to have not received the treatment (recall, this impact is not observable), $E(Y_i^0/D = 1, X)$, by re-weighting the sample of controls. Observe that the weights, $\omega(x_i)$, increase the relevance in the control sample of those individuals who are very similar to treated students, where similarity is defined here by the predicted probability of "participation" in a logit that explains participation given pre-treatment characteristics, that is, by the propensity score, $p(X_i)$. This allows us to compute the *inverse probability weighting estimator*. This estimator is achieved by regressing the outcome variable (either the PISA score or the probability of falling behind the lowest quartile) on the treatment, where each observation is weighted by $\omega(x_i)$. Thus, we are estimating a model for the outcome variable using the propensity score to weight our controls in the sample. By doing so, we obtain estimates of the average treatment effect for the treated (ATT), that is, the average effect for those students who attended PAE schools.[21] As there is a control for all covariates, X_i, in this estimation, through the consideration of the propensity score in the weighing procedure, there is no need to include them in the estimation. In any case, we may also include the covariates, X_i, in this regression as a robustness check.

Finally, we comment on the validity of the previous two assumptions. The second, the common support, can be tested by comparing the propensity score densities of the treated and control groups. We check this assumption graphically in the next section. However, the unconfoundedness assumption is difficult to validate. If it is not satisfied, this means that programme participation could be due, among other reasons, to special interest by parents, teachers or school principals. If these variables are positively correlated with the distribution of potential outcomes (i.e., more interested parents or teachers are also more likely to yield better student reading scores), then our estimates of the impact of the PAE would be biased; in particular, they would be overestimating the true impact of the programme. This assumption is therefore crucial. We attempt to address it by including a set of variables that might capture these parent, teacher and school principal characteristics (particularly whether parents exert pressure at the school and whether the principal is concerned with the school's reputation). In addition, in Section 5 below, we use the PISA 2009 scores to detect possible selection bias among schools participating in the PAE.

3.2 PAE participation

We estimate the predicted probability of participation in the remedial education programme (PAE) as a function of a set of characteristics of the students, parents and schools, i.e., the propensity score, $p(X_i)$. The set of variables included in X_i was chosen according to the differences in mean covariates in Table 6. We include indicators for female students, immigrant students, whether the student repeated a grade once or for more than one academic year, and whether the student attended pre-primary education. Regarding socioeconomic variables, we included whether the mother is highly educated and the index of educational materials at home. Finally, we also included a set of school characteristics, including its mean

[21] See Hospido et al. (2015) for a similar approach and Hirano et al. (2003) or Busso et at. (2014) for methodological details.

socioeconomic index value, the student-teacher ratio, its size, the proportion of dropouts, whether the school is above the 75th percentile in the distribution of the proportion of dropouts, whether it is a rural or urban school and whether the principal works to enhance the school's reputation. We then augment the basic logit model by including interactions that were statistically different from zero according to a two-sided t-test. The final specification is shown in Table 7. The first column presents the estimates of the propensity score for the *PAE-Immediacy* treatment. Columns 2 and 3 present the estimates of the propensity score for the *PAE-Intensity* treatment (PAE-Intensity 1-2 years and PAE-Intensity 3-4 years, respectively). As can be observed, the specifications of the three propensity scores are the same. This allows us to obtain comparable results across the different treatments.

Here Table 7: Propensity score estimation results

The estimates in the first column confirm the results of Table 6. Treated and control students are similarly likely to be girls. In addition, treated students are more likely to be immigrants and to have repeated at least one grade. However, once a complete set of control variables is considered, the mothers of treated and control students are similarly educated. The index of educational materials in the home also exhibits comparable values between treatments and controls. Regarding school variables, compared to control students, the schools of treated students are more likely to have a lower socioeconomic index value, a larger size, a larger proportion of dropouts, a lower teacher-student ratio, and principals who are less interested in enhancing the school's reputation. Finally, observe that the results of the propensity score for the three treatments are very similar, in particular regarding the school variables.

Figure 2 illustrates the densities of the predicted probabilities of participation in the PAE for the treated and control groups. Although the two distributions differ in form, the figure shows how similar the control and treatment samples are. First, the support of the values of the propensity score of treated students (solid line) and that of the control (dotted line) are the same: both range from 0 to approximately 0.8 (PAE-Immediacy), 0.6 (PAE-Intensity 1-2 years) or 0.8 (PAE-Intensity 3-4 years). Therefore, the common support assumption seems to hold in our sample. In addition, there is no concentration of predicted values around zero or one (which would mean that there are no comparable control students for some treated students).

Here Figure 2: Propensity score support

Finally, columns (4) and (5) of Table 6 present the means of the treated and control sample once the latter is re-weighted by $\omega(x_i) = \frac{1-\pi}{\pi} \times \frac{p(X_i)}{1-p(X_i)}$. First, although column (5) and column (1) should be exactly the same, as treated students receive a weight of 1, they do not coincide due to the existence of missing values in the weight variable (observe the lower number of observations in this column). The last column in Table 6 reports the differences in characteristics between treated and re-weighted controls. As can be observed, these are not statistically different from one another, particularly for the set of controls considered in the propensity score estimation (i.e., the balancing property is satisfied). Finally, note that the sample is also similar along characteristics that we do not include in the propensity score (class size, rural, etc.). An exception are the Student Admission and Staff Decision variables (the former being larger in the treatment group and the latter being larger in the control group). The similar composition of treated and re-weighted control groups even in characteristics omitted from the propensity score reinforces the credibility of the assumption that treated and re-weighted control students would have performed similarly had the treated students not been treated.[22]

4 Main results

In this section, we comment on the impact of the two treatments considered in our analysis: *PAE-Immediacy* and *PAE-Intensity*.

4.1 PAE-Immediacy and PAE-Intensity

The estimated effect of the *PAE-Immediacy* treatment is reported in the first column of Table 8. The first row in Panel A shows the re-weighting estimate without covariates (IPWEnc). Hence, this result can also be inferred from the first row in Table 6. The proportion of treated students in the first quartile in the reading score distribution is equal to 0.231, while that of the re-weighted control group is equal to 0.264. The -0.033 difference is the observed impact of the programme. The standard error accounts for arbitrary correlation at the school level and is equal to 0.019; thus, the estimate is only statistically significant at the 10% confidence level. The effect is quite similar when we include all of the variables considered in the logit model used to obtain the weights; specifically, it is equal to -0.030 and statistically significant at the 5% confidence level (IPWEwc, see row 2). The robustness of this result suggests that the specification of the model that predicts PAE participation is appropriate. Nevertheless, we go further and compare each treated student with her most similar associated control counterparts and thus provide results using several nearest neighbour propensity score estimators. In particular, we provide estimators by varying the number of nearest neighbours considered in the estimation from 2 to 8 (NNPS(2) to NNPS(6) in row 3 to row 6). As can be observed, the results are quite similar to those obtained by using the inverse probability weighting estimator. In particular, the larger the number of nearest neighbours used, the more similar the results are to the estimation without covariates. To summarize, we find that the probability of falling behind into the bottom part of the distribution is reduced by between 3% and 6% from receiving remedial education under the PAE.

Table 8: The impact of the PAE

Panel B in Table 8 shows the results regarding the effect of the programme on the mean reading score. Again, the result in the first row can also be inferred from the second row in Table 6. As can be observed, the estimate obtained when we do not include all of the

[22] See Lavy and Schlosser, 2005 or Hospido et al. 2015 for a similar test.

covariates is not statistically significant. Nevertheless, as commented above, the effect is much more precise when we hold constant all variables included in the conditional model. By doing so, we find that the estimated effect is equal to 5.53, which amounts to 6.4% of one standard deviation (=5.93/86.61). The point estimate when we use a nearest neighbour propensity score matching estimator is larger, 12.34, which equals 14.2% of one standard deviation.[23]

Table 9 illustrates how the PAE changes the overall distribution of reading scores. We present the estimated Cumulative Distribution Function (CDF) of the reading score for certain percentiles (see column 1 for the specific percentiles computed and column 2 for the corresponding value of the reading score distribution for the complete sample including all public schools). In columns 3-5 we present the values of the three CDFs: the CDF of reading scores among control students, the CDF of reading scores among re-weighted controls and the CDF of reading scores among the treated, for the PAE-Immediacy, PAE-Intensity 1-2 year and PAE-Intensity 3-4 year treatments in Panels A, B and C, respectively. Finally, in column 6, we present the difference between the last two (this column shows a rate equal to the CDF treated/CDF weighted controls minus one).

Table 9: Estimated CDF reading scores

As can be observed in the three panels, for each percentile, there is a lower fraction of students below that reading score among the control sample than among the treated sample. In addition, when comparing treated and re-weighted controls, we observed that the fraction of students below any score in the distribution among the treated sample is lower than among the re-weighted control sample (except for the PAE-Intensity 1-2 year treatment). As the re-weighted sample is, under our assumptions, the distribution of the scores that treated students would have achieved in the absence of the programme, that pattern suggests an overall increase in the distribution of reading scores. Finally, observe that the group of students who receive the larger impact from the PAE (both the PAE-Immediacy and the PAE-Intensity 3-4 year treatments) are those whose reading scores are between the 15th and 30th percentiles of the distribution, that is, precisely those students whose outcomes are among the main targets of the programme. Next, we analyse these students' performance in detail.

As previously noted, the results for the full sample presented above might not precisely capture the true impact of the PAE but merely its potential effect. On the one hand, we are assuming that all of the students in schools with the PAE are treated. However, some of them might not have received remedial education at all. Observe that by doing so, we are underestimating the impact of the PAE. On the other hand, by considering all of the

[23] The PAE had also a strong impact on maths and science outcomes. For example, according to the NNPS(2) estimator, students in schools than joined the programme at least during the 2011/12 year (PAE-Immediacy) significantly reduced their probability of belonging to the low-achievers group in maths and science by approximately 3.9% and 3.2%, respectively. It also improves their mean maths and science scores by approximately 9.7 and 7.8 PISA points, respectively. All estimates are statistically different from zero at the 5% confidence level or better. The complete results are not reported in the paper but are available from the authors upon request.

students in the PAE school as treated, we might well be capturing peer effects of treated on non-treated students. This assumption might induce an overestimation of the impact of the PAE on treated students. To argue that the effect analysed is closer to the actual effect of the intervention on treated students, we focus our main analysis on two sub-samples of our evaluation sample. First we consider students whose reading score is below the median value of the distribution.[24] This sub-sample consists of 5,427 individuals. By considering students with poor academic results, we increase the likelihood that they actually participated in the programme. Second, we consider students whose reading score is above the median value of the distribution. This sub-sample consists of 6,320 individuals. By considering students with high academic results, we reduce the likelihood that they actually participated in the programme and were subject to positive spillover effects from treated students.

Table 10 reports the main findings of this analysis. The first two columns provide results for the PAE-Immediacy treatment. Column 1 provides the results for the sub-sample of students below the median. It reports the impact on the probability of being in the first quartile and the impact on the mean reading score. The second column reports the results for the sub-sample of students above the median. It shows the probability of being above the third quartile and the impact on the mean reading score.

Table 10: The impact of the PAE: sub-samples

We find that the probability of falling behind into the bottom part of the distribution is reduced by approximately 4% to 5% for those students in the sub-sample below the median. Therefore, by considering the full sample of students at the school, we came close to estimating the true impact of the PAE on moving students out of low-achiever status, which is the main objective of the programme. We also find that, as expected, the programme had no effect on the probability of becoming a high achiever, that is, on the probability of belonging to the third quartile. We do not find evidence of spillover effects of potentially treated students on non-treated students (see Lavy and Schlosser, 2004 for a similar result). Finally, observe that the impact of the PAE on mean reading scores is smaller for both the sub-sample of students below and above the median than for students in the full sample. This might be due to the fact that by censoring the sample using the median reading score, we are not considering those cases of treated students who as a result of having received the PAE are above the median but who in the absence of the treatment would have remained below it.

The estimated effect of the *PAE-Intensity* treatment is reported in the second and third columns of Table 8. For those students in schools where the PAE was implemented for at most two of the last four years, the probability of falling behind into the bottom part of the distribution declines by (when statistically significant) between 2.8% and 3.9% relative to students in schools where the PAE was not implemented at all. However, for those students in schools where the PAE was implemented for at least three of the last four years, that

[24] The median for the PISA sample for all public schools (that is, including schools that might have participated in other remedial programmes) is 482.19.

probability declines by between 3.2% and 7.5%, relative to students in schools where the PAE was not implemented. Therefore, we can conclude that the PAE has an intensity effect: the larger the number of academic years for which it is implemented in a school, the more likely students are to leave the low-achievers' group. The bottom part of Table 8 reports the results regarding the possible intensity effect of the programme on the mean reading score. As can be observed, most estimates obtained for the effect of the 1-2 year treatment are not statistically significant. However, most estimates for the 3-4 year treatment are significant. Thus, we conclude that the PAE also has an intensity effect on mean reading scores. In particular, by receiving the PAE for at least three years, mean reading scores increase by between 10.7 and 16.2 PISA points, that is, between 12.3% and 18.7% of one standard deviation (10.7/86.61 and 16.2/86.61, respectively).

We next decompose the overall effect into the effect on the sub-samples of students below and above the median. These results can be found in Table 10, columns 3 to 6. We find that implementing the PAE for just one or two years has, if any, an impact on the sub-sample of students below the median. When significant, we find that it reduces the probability of falling behind into the bottom quartile among such students by approximately 4.6%. However, it has no impact on those students above the median: it does not significantly increase the probability of becoming part of the third quartile for these students. In addition, implementing the PAE for just 1 or 2 years has no effect on the reading scores of either sub-sample. However, implementing the programme for three or four years has an impact. In particular, the probability of falling behind into the bottom part of the distribution declines by between 4.2% and 4.6% for those students in the sub-sample below the median. In contrast to the results for the PAE-Immediacy treatment, we now find that the programme had an effect on the probability of becoming a high achiever, that is, on the probability of belonging to the third quartile. In addition, and similar to the PAE-Immediacy treatment, observe that the impact of implementing the PAE for three or four years on mean reading scores for both the sub-samples of students below and above the median is smaller than for students the full sample.

4.2 Heterogenous effects: rural vs. urban schools

As mentioned above, the PAE consisted of providing support (4 hours per week) to students with special needs and learning difficulties. This support was provided by after-school instructors or teachers from the student's own school who work with these students in small groups. These remedial classes were held during after-school hours (see Footnote 9 for additional details on programme implementation). Therefore, both teachers and students had to return to the school for the programme, which might be more difficult for teachers in urban schools than those in rural schools, as the former do not necessarily live close to the school. Therefore, we would expect gradual attrition in PAE participation among teachers in urban schools that, as a result, might reduce the effectiveness of the programme for students there. To assess whether there is heterogeneity in the impact of the PAE, we examine its impact on the previous reading outcomes by school type: rural or urban.

We define a rural school as one located in a community of fewer than 15,000 persons (i.e., a village or a small town) and an urban school as one located in a community of 15,000 or more persons (i.e., a town, city or large city). There are 220 urban schools (with 6,456 students) and 175 rural schools (with 4,669 students) in our sample. Table A.1 in Appendix 2 compares the characteristics of treated and control students in urban and rural schools. Students in urban and in rural schools differ in several dimensions. Reading outcomes (both the probability of belonging to the first quartile and reading scores) are better among students in urban schools than in rural schools. Moreover, the proportion of immigrant students is larger among urban schools. However, the proportion of students with an educated father or mother is lower among rural schools. In addition, the mean socioeconomic index exhibits much higher values among urban schools. Finally, urban schools are larger in size than rural schools.

The difference between treated and control students also differs between urban and rural schools. For instance, whereas control students in urban schools have better outcomes than treated students, the reverse occurs in rural schools. The distribution of socioeconomic characteristics also differs: in urban schools, parents of control students have higher schooling levels than their counterparts among treated students. Conversely, in rural schools, the proportion of educated parents (fathers) is larger among treated students than among controls.

To estimate the impact of the PAE on urban versus rural schools, we proceed as in the previous section. We first estimate the probability of participating in the PAE separately for students in urban and rural schools, considering individual, family and school characteristics, that is, the propensity score. Second, we use the estimated propensity score to construct the re-weighted sample of controls in urban and rural schools.[25] Finally, we use the previous results to compute the inverse probability weighting estimator (with and without covariates) and the nearest neighbour matching estimator. Table 11 compares the average outcomes of treated students in urban schools to students in rural schools. Panel A provides results for the impact of the programme on the probability of belonging to the lower achiever group (Reading25), and Panel B provides results for the impact on mean reading scores.

Table 11: PAE Impact: urban vs. rural schools

The results for the PAE-Immediacy treatment (columns 1 and 2 for students in urban and rural schools, respectively) indicate that the impact is much larger in rural schools. The probability of falling into the first quartile reduces by twice as much for students in rural schools than for students in urban schools (7.5% and 3.5%, respectively). The increase in mean reading scores is also larger among students from rural schools. The results for the PAE-Intensity treatment suggest several similar findings. First, again, the impact of the PAE on mean reading scores is larger for students in rural schools than students in rural schools, regardless of whether the school joined the programme for at most two years or more than two years. Second, similar to the results for the full evaluation sample, the PAE has an

[25] Columns (3) and (6) in Table A.1 show that treated and control samples, in both urban and rural schools, are comparable once re-weighted.

intensity effect in rural schools: the larger the number of academic years it is implemented in a school, the higher the probability of students leaving the low-achievers' group and the higher the increase in mean reading achievement. However, the PAE has no intensity effect for students in urban schools.

To conclude, the impact of the PAE is, in general, much larger for students in rural than in urban schools. The information reported above regarding the implementation of the programme might provide a possible explanation for the sources of these different results for students in urban versus rural schools, without attaching any causal interpretation: namely, offering remedial classes in after-school hours might be more difficult to implement for teachers in urban schools. As a result, it could be the case that some of them do not teach the total number of remedial classes or even abandon the programme.

5 Selection bias: are PAE schools different from the rest?

As previously noted, our results above can be called into question based on the argument that treated schools volunteer for the programme, while control schools did not. Therefore, it is possible that principals who decide to participate in the PAE have unobserved characteristics that correlate with students' characteristics and with their outcomes. Similarly, students in treated schools may have unobserved characteristics that correlate with the decision of the principals to join the PAE and with reading scores. If these unobserved school (principal, teacher, etc.) characteristics are positively correlated with students' outcomes, then our previous results would be overestimating the true impact of the programme. For example, highly motivated and active principals may, in addition to deciding to participate in the PAE, promote various types of activities and initiatives to improve their students' results. However, these unobserved school (principal, teacher, etc.) characteristics might also be negatively correlated with students' outcomes, for example, the existence of a difficult student body at the school. In that case, then our previous results would be underestimating the true impact of the programme. Thus, it is very difficult to establish a priori the sign and magnitude of the bias. Formally, according to Heckman et al. (1998), we can define selection bias as follows. Let first consider the linear model:

$$Y_t = X_t'\beta + u_t,$$

where Y_t is the student's outcome at time t and X_t is a set of observables (individual, family and school variables). Now, suppose that one of the school characteristics is PAE participation. Then, the conditional average of the Y variable given specific values of the regressors would be calculated as follows:

$$E(Y_t \mid X_t = x, PAE_t = pae) = x'\beta + \delta pae \tag{7}$$

However, as noted previously, it is very difficult to conclude that parameter δ is capturing the impact of PAE participation due to possible selection bias. We partially addressed this problem by including in the covariates a set of variables capturing principal characteristics that might be both affecting students' scores and the probability of participating in the PAE. Here, we use the PISA 2009 dataset to characterize possible selection bias under the assumption that the true impact of a non-existent programme is zero. In particular, we replicate the analysis in (7) by replacing PAE_t with an indicator, D_{t+1}, indicating that the school participated in the PAE after the 2008/09 academic year but not before that date:

$$E(Y_t \mid X_t = x, D_{t+1} = D) = x'\beta + \alpha D \tag{8}$$

Observe that the student outcome, Y_t, is measured at time t (in this case, the PISA 2009 scores), whereas the treatment, D_{t+1}, is measured at time $t+1$, as it will occur well after the PISA 2009 scores were measured (treated schools will be those that did not participate in the PAE between the 2005/06 and 2008/09 academic years, but did participate thereafter) Thus, if there is no selection bias, the estimated impact of this "treatment" should be zero. We next estimate its impact following the empirical strategy presented in a previous section. We first estimate the predicted probability of participating in the PAE only after the 2008/09 academic year by considering a set of individual, family and school variables. Then, we re-weight the control group such that their observable re-weighted characteristics are statistically similar to those of the treatment group. Finally, we estimate the (non-existent) effect of participating in PAE after the 2008/09 academic year for the treated students (using PISA 2009 scores). As above, this allows us to compute the inverse probability weighting estimator (IPWE). This is achieved by regressing the outcome variable (either the PISA 2009 score or the probability of falling into the lowest quartile) on the "treatment", where each observation is weighted by $w(x_i)$. We also include the covariates, X_i, in the regression as a robustness check. In addition, we compute the nearest neighbour propensity score (NNPS) estimators after verifying that our estimates of the propensity score fulfil the balancing property. We finally compare results from following the two empirical strategies.

5.1 The data

Our sample now consists of 4,568 students from 144 schools that participated in both PISA 2009 and PISA 2012.[26] Therefore, for those schools, we know whether they participated in the PAE in any academic year since the programme began. In particular, 31 such schools participated in the programme only after the 2008/09 academic year. Thus, the sample consists of 912 "treated" students and 3,656 "control" students.[27] Using the rich information from the PISA 2009 database, we can compare them according to individual, parental and school variables. In addition, we can also identify which variables account for the possible selection bias. Table 12 below compares the characteristics of treated and control students

[26] We also exclude, as in previous exercises, schools that participated in other remedial education programmes between the 2005/06 and 2011/12 academic years.

[27] We believe that our sample size is large enough to provide relevant results. See, for example, Heckman et al. (1998) who use samples of approximately 200 treated subjects to study the properties of selection biases in employment programmes.

in the sample.

Here Table 12: Summary statistics: treated and control. Selection bias

Although there are no significant differences regarding gender composition between the two groups, students in schools that subsequently participated in the PAE differ from controls in an important number of characteristics. As can be observed in Table 11, treated students are more likely to be immigrants and are 10 points more likely to have repeated a grade at least once. In addition, the proportion of educated parents (mother and/or father), the index of educational materials and the mean socioeconomic index are lower among treated students, suggesting that treated schools have a higher proportion of students from disadvantaged backgrounds. Finally, treated students came from smaller sized schools where the proportion of educated parents is lower than that for controls. Conversely, students in the control sample are from schools with a larger student-teacher ratio. Finally, treated students performed worse on PISA 2009: the proportion of students with a reading score in the first quartile is larger among treated students, and their mean reading score is lower than among controls. Therefore, the previous results suggest that, if any selection into participation in the PAE based on unobservable characteristics exits, then these variables are negatively correlated with students' outcomes, which implies that our previous estimates are underestimating the true impact of the programme.

5.2 Selection bias estimation

As we know that PAE participation in treated schools occurred well after the PISA 2009 exams took place, the difference in reading outcomes (once we control for student, parent and school characteristics) can only be due to the influence of unobserved variables or selection bias. Next, we examine the possible existence of selection bias. We first estimate the probability of participating in the PAE only after the 2008/09 academic year. We present the results in Table 13.

Here Table 13: Determinants of PAE participation only after 2009

The analysis of participation determinants confirms that the treated group contains a larger proportion of immigrants and repeater students. In addition, having an educated mother (and living outside Basque Country) or a large index of educational materials reduces the probability of participation. Similarly, the results in Table 13 suggest that the schools of the treated students are more likely to have a lower socioeconomic index value, be of a smaller size and are also more likely to be located in urban municipalities. Column (3) of Table 12 shows the average characteristics of the control group once it is re-weighted according to the predicted probability of participation. Observe that, again, the number of observations for re-weighted controls is reduced due to the existence of missing values for the weighting variable.[28] It can be seen that the sample of control students, once re-weighted, is similar to

[28]Summary statistics for the sample of treated students for which the weighting variable is not missing are

that of the treated students in terms of reading outcomes and individual, family and school variables.

Here Table 14: Impact of PAE participation (only after 2009)

Finally, we proceed to estimate the impact of PAE participation only after 2009. The upper part of Table 14 reports the results for the probability of belonging to the first quartile of the reading score distribution. The first row shows the result of a simple probit estimation. As can be observed, the effect of programme participation after 2009 is zero. The second row shows the re-weighting estimate without covariates. Therefore, this result can also be inferred from the first row of Table 12. The 0.002 difference is the observed impact of the programme (see Footnote 25). The standard error accounts for arbitrary correlation at the school level and is equal to 0.0042; thus the estimate is also not significantly different from zero. The effect is quite similar when we include all variables considered in the logit model used to obtain the weights (third row). In addition, we go further and compare each treated student with her most similar associated control counterparts and thus provide results using two nearest neighbour propensity score estimators. As can be observed, the results are remarkably similar to those obtained using the inverse probability weighting estimator.

The bottom part of Table 14 shows the results regarding the effect of the programme on the mean reading score. Again, the results in the second row can also be inferred from the second row in Table 12. The estimate obtained when we do not include all of the covariates is not statistically significant. Nevertheless, as noted above, the effect is much more precise when we hold constant all of the variables included in the logit model. By doing so, we also find that the estimated effect is not statistically different from zero. The point estimate when we use the nearest neighbour propensity score matching estimator is also negligible, at 2.915, and is not statistically different from zero.

To conclude, we find that the results of the schools that participated in the PAE only after the 2008/09 academic course were not very different from the rest, suggesting that no selection bias exists. Nevertheless, if any, possible differences can be explained by differences in individual, parental and school characteristics. Accounting for these differences completely attenuates the selection bias.[29] Therefore, our results above suggest that it is feasible to obtain estimates of the impact of PAE participation on reading outcomes with no selection bias by re-weighting the sample according to student, family and school characteristics, as we have done above. A possible explanation for the lack of selection is that, as the programme was introduced in the 2005/06 academic year, by the 2009/10 academic year, and thereafter, the existence of the programme was sufficiently widespread in the education community (indeed, the rate of participation in the programme exceeded 45% in some regions).

not reported here for clarity but are available upon request.

[29]As an additional robustness check, note that neither the 95% confidence interval for the estimated bias for the probability of being in the first quartile of the reading distribution [-0.0583,0.0541] (Table 13, row 4) nor the 95% confidence interval for the estimated bias for the mean reading score [-5.426, 11.257] (Table 13, row 8) contains the point estimate of the effect of the PAE in Table 8 above.

6 Robustness analysis

Finally, we want to check whether our previous results when considering the student as the unit of analysis hold when we instead consider the school as the unit of analysis. Recall that to the extent that we cannot observe whether a particular student actually received the treatment, our previous findings merely suggest the *potential* effect of the PAE. By considering the school as the unit of analysis, and similar to Lavy and Schlosser (2005), two problems emerge. First, if there is a small number of treated students at a school it may be very difficult to observe any effect. In addition, to claim that the effect analysed is the actual or true effect of the intervention on treated students, we need to assume that the PAE did not generate spillover effects on non-treated students (which appears to be the case in light of our previous results from the sub-sample of students with reading scores above the median).

Table A.2 in Appendix 2 provides the summary statistics of the schools in our evaluation sample and for all public schools. Table A.3 compares the characteristics of treated and control schools, which differ in several dimensions: first, the mean reading score is higher among treated schools. In addition, the proportions of repeaters, immigrants and dropouts are also larger among treated schools. Conversely, the proportion of educated parents and the socioeconomic index is higher among the control group. Furthermore, the proportion of treated schools where the principal claims that he/she works to enhance the school's reputation is nearly twice as large relative to control schools.

To estimate the impact of the PAE on schools, we proceed as above. We first estimate the probability of participating in the PAE considering only school characteristics, that is, the propensity score.[30] Second, we use the estimated propensity score to construct the re-weighted sample of control schools.[31] Finally, we use the previous results to compute the inverse probability weighting estimator (with and without covariates) and the nearest neighbour matching. Table 15 provides the estimated effect of the PAE-Immediacy treatment (column 1) and the PAE-Intensity treatment (columns 2 and 3). Panel A provides results for the impact of the programme on the probability of belonging to the lower-achiever group (Reading25), and Panel B provides results for the impact on mean reading scores.

Here Table 15: The impact of the PAE: schools

As can be observed, the results are very similar to those found in Table 8 when considering the student as the unit of analysis. In particular, the impact of the programme now appears to be larger. First, the proportion of students at the school in the first quartile of the distribution declines by between 4.9% and 7.6%, depending on the estimator (compared to the 3%- 6% reduction at the student level), in those schools that participated in the PAE at least during the 2011/12 academic year (PAE-Immediacy treatment). The results regarding the effect of the programme on the mean reading score at both the school and the student level

[30] Table A.4 in Appendix 2 presents the results for the estimated propensity score for the PAE-Immediacy and PAE-Intensity treatments.

[31] Column (3) in Table A.3 indicates that treated and control schools are comparable once re-weighted.

are not significantly different when we use a nearest neighbour propensity score matching estimator, approximately 12.13 PISA points (12.34 in Table 8 above). Nevertheless, the inverse probability weighting estimator produces larger impacts at the school level than at the student level in this case. Finally, we also find that the PAE has an intensity effect: the larger the number of academic years for which it is implemented at the school, the larger the proportion of students exiting the low-achiever group. The programme has almost no impact in those schools that participated in the programme for at most two academic years, whereas it has a strong impact among those schools that participated for at least three years: the proportion of students in the low-achiever group declines by between 5% and 8%, and the mean reading score increases by between 10.3 and 19.8 PISA points.

7 Concluding remarks

There is ample evidence of increasing inequality and poverty figures in developed countries. As a result, addressing early school leaving and improving the education and skills of the workforce are priorities of policy makers in several countries. National governments are currently being encouraged to undertake evidence-based education policies to reduce the adverse effects of the aforementioned facts. Surprisingly, it is difficult to find empirical evidence regarding the effectiveness of most of these interventions and in particular remedial education programmes. In this paper, we estimate the effects of a remedial programme implemented in Spain between 2005 and 2012 that offered additional instruction time for underperforming students from poor socioeconomic backgrounds: the Programme for School Guidance (PAE). Our main finding is that this programme had a substantial positive effect on students' academic achievement. First, our results suggest that it reduced the probability of falling behind into the bottom of the reading score distribution by approximately 5% (nearly 10% of one standard deviation). The estimated effect on mean reading scores is above 12 PISA points (more than 14% of one standard deviation). We also find that a larger exposure to the programme improves students' scores. Furthermore, our evidence suggests that there is heterogeneity in the impact of the programme across types of schools, urban versus rural, with the impact being much larger among students attending rural schools than urban schools.

This study has an important limitation. Namely, we lack data on whether a particular student actually received the treatment and instead merely observe whether the student attended a school that participated in the programme, implying that the effects we obtain can only be understood as the *potential* effects of the programme. We address this shortcoming by performing two additional tests. We first decompose our evaluation sample into two sub-samples: one with students whose reading scores are below the median value of the distribution, that is, students who were more likely to received the treatment, and other with students whose reading scores are above the median, that is, students who might not have participated in the programme but received positive spillover effects from treated students. In addition, we check whether our previous results hold when we instead consider the school as the unit of analysis. By proceeding with these strategies, we conclude that our previous results are, if anything, underestimating the true impact of the programme on treated students.

Future research should proceed by evaluating the impact of this or similar programmes on a wider range of student outcomes, such as dropout, absenteeism or even on non-cognitive skills, such as study habits (motivation and discipline), self-esteem, and confidence (see, among many others, Heckman et al., 2006 on the growing literature demonstrating that young students' non-cognitive skills significantly affect their school achievement and work outcomes). In this study, we examine only short-term effects due to a lack of sufficient data on schools participating in the programme only well before the PISA 2012 exams. Nevertheless, learning about the long-run effects of the programme is required to fully understand its effectiveness.

We believe that our results are of value and contribute novel, interesting insights to a relatively scarce literature on remedial education programmes and their impact on under-performing teenagers across Europe. In this paper, we find support for policies consisting of targeted additional instruction time to improve poor-performing students' achievement.

References

[1] Atkinson A. (2010): "Macerata lectures on European economic policy. Poverty and the EU: the new decade," *Working Paper* 24-2010, Macerata University, Department of Studies on Economic Development (DiSSE).

[2] Atkinson A., Piketty T., and E. Saez (2011): "Top incomes in the long run of history" *Journal of Economic Literature* 49, 1: 3-71.

[3] Banerjee, A. V., S. Cole, E. Dufflo and L. Linden (2007): "Remedying Education: Evidence from Two Randomized Experiments in India" *The Quarterly Journal of Economics* 122 (3):1235-1264.

[4] Battaglia, M. and L. Lebedinski (2015): "Equal Access to Education: An Evaluation of the Roma Teaching Assistant Program in Serbia", *World Development*, 76: 62-81

[5] Bettinger E. and B. Long (2009): "Addressing the needs of under-prepared college students: does college remediation work?" *Journal of Human Resources* 44: 736–771

[6] Brunello, G. and M. De Paola (2014): "The Costs of Early School Leaving in Europe" IZA Journal of Public Policy 3(22).

[7] Busso, M., J. DiNardo, and J. McCrary (2014): "New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators" *Review of Economics and Statistics* 96 (5): 885–897.

[8] Calcagno J.C. and B. T. Long (2008): "The Impact of Postsecondary Remediation Using a Regression Discontinuity Approach: Addressing Endogenous Sorting and Noncompliance" NBER Working Paper No. 14194

[9] De Paola, M. and V. Scoppa (2014): "The Effectiveness of Remedial Courses in Italy: A Fuzzy Regression Discontinuity Design" *Journal of Population Economics*, 27(2): 365-386

[10] De Paola, M. and V. Scoppa (2015): "Procrastination, Academic Success and the Effectiveness of a Remedial Program" *Journal of Economic Behavior and Organization*, 115: 217–236

[11] European Commission (2013a). Social Investment Package. Brussels.

[12] European Commission (2013b). Education and Training in Europe 2020: Responses from the EU Member States. Eurydice Report. Brussels: Eurydice

[13] Freeman, R. (2008). "Globalization and Inequality", in the Oxford Handbook of Economic Inequality, edited by W. Salverda, B. Nolan and T. Smeeding. Oxford: Oxford University Press.

[14] García-Pérez, J.I., Hidalgo-Hidalgo M. and J. A. Robles-Zurita (2014): "Does grade retention affect students' achievement? Some evidence from Spain" *Applied Economics* 46 (12): 1373–1392.

[15] Jacob, B. A. and L. Lefgren (2004): "Remedial Education and Student Achievement: A Regression Discontinuity Analysis" *Review of Economics and Statistics* 86 (1): 226-244.

[16] Heckman, J. J., Ichimura, H., Smith, J. and Todd, P. (1998): "Characterizing Selection bias using Experimental Data" *Econometrica* 66(5): 1017-1098.

[17] Heckman, J., J. Stixrud and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labour Market Outcomes and Social Behavior" *Journal of Labor Economics* 24(3): 411-480.

[18] Hirano, K., G. Imbens and G. Ridder (2003) "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score" *Econometrica* 71(4): 1161-1189.

[19] Holmlund, H. and O. Silva (2014): "Targeting Non-Cognitive Skills to Improve Cognitive Outcomes: Evidence from a Remedial Education Intervention" *Journal of Human Capital* 8(2): 126-160.

[20] Hospido, L., E. Villanueva and G. Zamarro (2015): "Finance for All : The Impact of Financial Literacy Training in Compulsory Secondary Education in Spain" Banco de España WP 1502 2015, IZA DP 8902 2015

[21] Kanbur, R. (2014): "Globalization and Inequality" Working Papers 180163, Cornell University, Department of Applied Economics and Management.

[22] Kremer, M., C. Brannen and R. Glennerster (2013): "The Challenge of Education and Learning in the Developing World" *Science* 340 : 6130 (April 19): 297-300.

[23] Lavy, V. and A. Schlosser (2005): "Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits" *Journal of Labor Economics* 23 (4): 839-874.

[24] Myoung-Jae Lee (2005) "Micro-Econometrics for Policy, Program, and Treatment Effects", Advanced Texts in Econometrics (Ed. Ganger, C. W. J. y Mizon, G. E.), Oxford University Press.

[25] OECD (2011) PISA 2009. Technical Report OECD Publishing, OECD, Paris.

[26] OECD (2013). Crisis squeezes income and puts pressure on inequality and poverty. OECD Publishing, Paris.

[27] Rosenbaum, P. R. and Rubin, D. B. (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects" *Biometrika* (70): 41-75.

[28] Rosenbaum, P. R. and Rubin, D. B. (1984): "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score" *Journal of the American Statistical Association* 79: 516-524.

[29] Spanish Ministry of Education (2016): Estadísticas de las Enseñanzas no universitarias, Ministerio de Educación Cultura y Deporte. Spain. Available at http://www.mecd.gob.es/servicios-al-ciudadano-mecd/estadisticas/educacion/no-universitaria/centros/centros-servicios-estadisticas.html (accessed 15 July 2016).

[30] U.S: Department of Education, Office of the Under Secretary (2003): When schools stay open late: the national evaluation of the Twenty-First Century Community Learning Centers Program, first year findings. U.S. Department of Education, Office of the Under Secretary, Washington, DC.

Appendix 1: Variable Description

We describe all of the variables used in our estimations (the original variable names in the PISA database are presented in capital letters).

- PAE-Immediacy: Dummy variable that equals 1 for students at schools that participated in the PAE during the 2011/12 academic year (0 for students at schools that never participated in the PAE). Source: INEE.
- PAE-Intensity 1-2 years: Dummy variable that equals 1 for students at schools that participated in the PAE during the 2010/11 and/or 2011/12 academic years (0 for students at schools that never participated in the PAE). Source: INEE.
- PAE-Intensity 3-4 years: Dummy variable that equals 1 for students at schools that participated in the PAE for 3 out of 4 of the academic years between 2008/09 and 2011/12 (0 for students at schools that never participated in the PAE). Source: INEE.
- Reading: The average of the five plausible values of literacy outcomes. Source: Students questionnaire PISA 2012.
- Reading25: Dummy variable equal to 1 for students with reading score below the first quartile of the reading scores distribution for all public schools, i.e., 420.1. Source: Students questionnaire PISA 2012.
- Immigrant: Dummy variable equal to 1 for non-native students, i.e., first- or second-generation immigrants (IMMIG above 1). Source: Students questionnaire PISA 2012.
- Repeater once: Dummy variable equal to 1 for students attending grade 9 (ST01Q01 equal to 9). Source: Students questionnaire PISA 2012.
- Repeater more once: Dummy variable equal to 1 for students attending grade 8 or lower (ST01Q01 lower or equal to 8). Source: Students questionnaire PISA 2012.
- Attended pre-primary: Source: Dummy variable equal to 1 for students attending pre-primary schools for more than one year (ST05Q01 above 2). Source: Students questionnaire PISA 2012.
- Mother/father educated: Dummy variable equal to 1 for students whose mother/father attended at least tertiary education. Source: Parents questionnaire PISA 2012.
- Index educational materials: Index of whether the home possesses a desk and a quiet place to study, a computer and/or educational software, books to help with schoolwork and a dictionary (HEDRES variable ranges from -3.93 to 9999.00 in the international dataset). Source: Parents questionnaire PISA 2012.
- Students educ parents: % Students at school with educ. parents: Percentage of students at the school whose mother and father attended at least tertiary education. Source: Parents questionnaire PISA 2012.

- ESCS: Dummy variable equal to 1 if the school is above the last third in the index of economic, social and cultural status distribution for all public schools. Source: Parents questionnaire PISA 2012.
- School size: Total school enrolment. Source: School questionnaire PISA 2012.
- Presion: Perceptions of principals about parents exerting pressure towards the school to set high academic standards and to have their students achieve them (SC24Q01 below 3, which is the answer to what best characterizes parental expectations towards your school. Possible answers range from 1-there is constant pressure- to 3 -pressure largely absent). Source: School questionnaire PISA 2012.
- Proportion of Dropout students at School: Proportion of students who left the school without the certificate that allows them to enter post-secondary or vocational education, apprenticeships or employment (generated from 2-digit variable SC23Q01). Source: School questionnaire PISA 2012.
- Proportion of Dropout students (percentile 75): Dummy variable equal to 1 if the school is above the 75th percentile in the evaluation sample in the proportion of dropout students at school.
- Student Teacher Ratio at School: number of students per teacher at the school (generated from 2-digit variable STRATIO). Source: School questionnaire PISA 2012.
- Rural School: Dummy variable equal to 1 if the school is located in a village or a small town and to 0 if located in a town, a city or large city (SC03Q01 is the principal answer to school location). Source: School questionnaire PISA 2012.
- Principal Enhances Reputation: Dummy variable equal to 1 if the principal says that he/she works to enhance the school's reputation in the community once or more than once per week (SC34Q01 above 5). Source: School questionnaire PISA 2012.
- Students Admittance: Dummy variable equal to 1 if the principal says that the school has responsibility for the student's admittance (SC33Q09C equal to 1). Source: School questionnaire PISA 2012.
- Staff decision: Dummy variable equal to 1 if the principal says that the school has responsibility for staff hiring decisions (SC34Q10 between 4 and 6 both included). Source: School questionnaire PISA 2012.
- Review Work: Dummy variable equal to 1 if the principal says that he/she reviews work produced by students when evaluating classroom instruction (SC34Q20 above 5). Source: School questionnaire PISA 2012.
- Discuss Problems: Dummy variable equal to 1 if the principal says that he/she takes the initiative to discuss matters when the teacher has a problem in the classroom (SC34Q07 above 5). Source: School questionnaire PISA 2012.

- Assess: Dummy variable equal to 1 if the "Use of Assessment" index (ASSESS) is equal to 5 or 6 in the school. This index measures the extent to which assessments of students are used to inform parents of their child's progress, to make decisions about students' retention or promotion, to group students for instructional purposes, to compare the school to district or national performance, etc. Source: School questionnaire PISA 2012.

Table 1: PAE program implementation

Academic courses					Schools		Reading (mean)	ESCS (mean)
					Number	%		
2005-2008	2008-2009	2009-2010	2010-2011	2011-2012				
X	X	X	X	X	45	0.108	473.113	-0.322
-	X	X	X	X	15	0.036	465.913	-0.666
-	-	X	X	X	29	0.070	474.023	-0.671
X	-	X	X	X	0	0	-	-
X	X	-	X	X	4	0.010	372.100	-0.995
X	-	-	X	X	0	0	-	-
-	X	-	X	X	0	0	-	-
-	-	-	-	X	17	0.041	491.868	-0.492
-	-	-	X	X	19	0.046	473.278	-0.321
X	X	X	-	X	0	0	-	-
X	X	-	-	X	0	0	-	-
X	-	X	-	X	0	0	-	-
X	-	-	-	X	0	0	-	-
X	X	X	X	-	7	0.017	484.040	-0.136
X	X	-	X	-	1	0.002	277.615	-1.160
X	-	X	X	-	0	0	-	-
-	X	X	X	-	0	0	-	-
X	-	-	X	-	0	0	-	-
-	X	-	X	-	0	0	-	-
-	-	X	X	-	6	0.014	471.557	-0.030
-	-	-	X	-	0	0	-	-
-	-	-	-	-	266	0.638	476.450	-0.229
X	X	X	-	-	1	0.002	532.874	0.810
X	X	-	-	-	5	0.012	411.166	-1.308
X	-	X	-	-	0	0	-	-
-	X	X	-	-	1	0.002	349.148	-1.510
X	-	-	-	-	0	0	-	-
-	X	-	-	-	1	0.002	512.812	-1.110
-	-	X	-	-	0	0	-	-
					417	1.000	473.74	-0.321

Note: X(respectively, -) indicates the school participated (respectively, did not participate) in the PAE program in the corresponding academic course. Source: INEE (National Institute for Educational Evaluation) and PISA 2012.

Table 2: Schools with PAE in PISA 2012

		2005/2006		2006/2007		2007/2008		2008/2009		2009/2010		2010/2011		2011/2012	
	PISA 2012	PAE	PISA	PAE	PISA	PAE	PISA	PAE	PISA	PAE	PISA	PAE	PISA	PAE	PISA
Andalusia	52	37	4	72	3	161	9	200	11	320	16	350	16	400	7
Aragon	51	4	1	7	3	15	3	19	3	28	6	31	8	50	16
Asturias	56	3	2	5	2	11	3	11	5	11	6	11	6	11	6
Balearic Islands	54	0	0	0	0	0	0	10	6	15	10	15	14	26	16
Cantabria	54	2	2	4	4	8	7	10	9	10	5	18	6	19	6
Castile Leon	55	8	3	15	5	33	5	36	6	36	7	36	7	36	7
Catalonia	51	20	0	36	0	71	4	71	5	92	4	92	4	92	4
Extremadura	53	6	2	11	2	23	4	23	4	37	8	50	11	54	15
Galicia	56	10	1	19	2	40	8	40	8	45	4	45	8	49	10
La Rioja	54	1	1	5	5	10	9	13	10	12	15	15	19	17	19
Madrid	51	11	1	26	2	78	6	100	9	109	11	114	11	126	14
Murcia	52	6	1	11	5	26	11	28	10	39	14	51	19	51	18
Navarre	51	1	0	3	1	6	3	6	3	7	2	8	2	10	2
Basque Country	174	0	0	4	3	11	2	13	3	30	13	42	20	50	23
TOTAL	902	149	20	289	39	587	76	692	97	908	130	984	154	1102	165

Source: INEE (National Institute for Educational Evaluation) and PISA 2012

Table 3: Summary Statistics

	Evaluation sample	All public schools
Reading scores		
Mean	481.0	476.6
Standard Deviation	86.61	88.58
Individual Variables		
Gender (girl)	0.499	0.497
Immigrant	0.107	0.116
Repeater once	0.261	0.270
Repeater more	0.119	0.128
Attended pre-primary	0.824	0.824
Socioeco background		
Father educated	0.317	0.308
Mother educated	0.309	0.301
Index educ possesions	0.068	0.047
School Variables		
Students educ parents	0.179	0.172
ESCS	-0.322	-0.369
Presion	0.339	0.331
School size	594.2	595.8
Prop Immigrants	0.105	0.113
Prop Dropout	0.096	0.102
Student Teacher Ratio	10.36	10.11
Rural	0.386	0.364
Ppal Enhance Reputat.	0.252	0.255
Observations	11,747	15,296

Note: Evaluation sample: PISA sample excluding students in private schools and schools which joined other remedial programs. See the text for details. Source: PISA 2012

Table 4: Treatments definitions

Academic courses					PAE-Treatments		
					PAE-Immediacy	PAE Intensity	
2005-2008	2008-2009	2009-2010	2010-2011	2011-2012		1-2 Years	3-4 years
X	X	X	X	X	1	.	1
-	X	X	X	X	1	.	1
-	-	X	X	X	1	.	1
X	X	-	X	X	1	.	1
-	-	-	-	X	1	1	.
-	-	-	X	X	1	1	.
X	X	X	X	-	.	.	1
X	X	-	X	-	.	1	.
-	-	X	X	-	.	1	.
-	-	-	-	-	0	0	0
X	X	X	-	-	.	1	.
X	X	-	-	-	.	1	.
-	X	X	-	-	.	1	.
-	X	-	-	-	.	1	.

Note: X(respectively, -) indicates the school participated (respectively, did not participate) in the PAE program in the corresponding academic course. 1(resp., 0) indicates whether the schools participating in PAE in the academic courses shown in that row with an X are consider as treated (resp., controls) according to the different treatment definitions. · means that these schools are dropped from the analysis.

Table 5: Treated and control: schools and students

PAE-Treatments		Treated		Control	
		Schools	Students	Schools	Students
PAE Immediacy		129	3,666	266	7,459
PAE Intensity	1-2 Years	51	1,425	266	7,459
	3-4 Years	100	2,863	266	7,459

Note: The number of schools and students corresponds to the Evaluation Sample. Source: PISA 2012

Table 6: Summary Statistics: Treated and control

	Treated (1)	Controls (2)	Diff (1)-(2)	Treated[i] (4)	Weighted Controls (5)	Diff (4)-(5)	Pscore
Reading Scores							
Reading25	0.234	0.215	0.019**	0.231	0.264	-0.033***	
Reading	479.9	487.2	-7.300***	480.9	474.7	6.200***	
Individual variables							
Gender (girl)	0.499	0.508	-0.009	0.501	0.497	0.004	yes
Immigrant	0.155	0.09	0.067***	0.151	0.156	-0.005	yes
Repeater once	0.271	0.225	0.046***	0.269	0.269	0.000	yes
Repeater more once	0.130	0.09	0.041***	0.128	0.133	-0.005	yes
Attended pre-primary	0.813	0.829	-0.016***	0.818	0.819	-0.001	yes
Socioeconomic Variables							
Father educated	0.297	0.346	-0.049***	0.299	0.302	-0.003	no
Mother educated	0.301	0.366	-0.065***	0.303	0.311	-0.008	yes
Index of educ pos	0.041	0.07	-0.033*	0.0408	0.058	-0.017	yes
School variables							
Stu Teacher Ratio	8.55	9.134	-0.589***	8.548	0.295	8.253	yes
ESCS	0.284	0.445	-0.161***	0.285	0.007	0.278	yes
School size	589.10	557.70	31.400***	589.70	593.40	-3.700	yes
Ppral Enhance repu	0.216	0.236	-0.020**	0.216	0.213	0.003	yes
Prop Dropouts	0.12	0.09	0.031***	0.116	0.119	-0.003	yes
Dropout75	0.294	0.221	0.073***	0.293	0.303	-0.010	yes
Stud Admin	0.394	0.336	0.058***	0.394	0.279	0.115***	no
Staff Dec	0.629	0.796	-0.167***	0.630	0.826	-0.196***	no
Review Work	0.167	0.147	0.020***	0.167	0.179	-0.012	no
Discuss Problems	0.311	0.332	-0.021**	0.311	0.325	-0.014	no
Asses	0.413	0.446	-0.033***	0.413	0.416	-0.003	no
Rural	0.405	0.427	-0.022**	0.406	0.415	-0.009	no
Classize	21.44	21.67	-0.230	21.43	21.57	-0.140	no
Observations	3,666	7,459		3,630	7,063	7,395	

Note: Treated students under Treatment PAE-Immediacy. Treated and control in columns (1) and (2) are sample averages. Treated and controls in columns (4) and (5) are averages of the treated and control group when the sample is reweighted by the (inverse of the) probability of receiving PAE-Immediacy treatment predicted by the set of individual, socioeconomic and school variables in this table. (i) column (4) and column (1) should be exactly the same as treated students receive a weight of 1, however they do not coincide due to the existence of missing values in the weight variable (observe the reduced number of observations in columns (4) and (5). We test for mean differences: *** p<0.01, ** p<0.05, * p<0.1. Source: PISA 2012

Table 7: Propensity score estimation

	PAE-Immediacy	PAE-Intensity 1-2 years	PAE-Intensity 3-4 years
Individual			
Gender (girl)	-0.019	-0.036	-0.000
	(0.044)	(0.061)	(0.047)
Immigrant	0.423***	0.101	0.522***
	(0.075)	(0.109)	(0.079)
Immigrant x Murcia	-0.625***	-0.437	-0.726***
	(0.239)	(0.436)	(0.253)
Immigrant x Extrem	-0.239	0.169	-0.225
	(0.512)	(0.808)	(0.571)
Repeater once	0.152***	0.192***	0.250***
	(0.053)	(0.073)	(0.057)
Repeater more once	0.183**	0.249**	0.285***
	(0.075)	(0.105)	(0.080)
Attended pre-primary	-0.026	0.030	0.032
	(0.062)	(0.085)	(0.066)
Socioeconomics			
Mother educated	-0.063	-0.208***	-0.014
	(0.053)	(0.077)	(0.056)
Mo educ x BasqueCountry	-0.135	0.495***	-0.471***
	(0.098)	(0.132)	(0.119)
Index educ pos	0.008	0.025	0.020
	(0.025)	(0.035)	(0.028)
School variables			
Stu Teacher Ratio	-0.082***	-0.096***	-0.123***
	(0.012)	(0.018)	(0.013)
ESCS	-0.908***	-1.087***	-0.756***
	(0.051)	(0.079)	(0.055)
Rural x Anda	-1.371***	0.673***	0.301*
	(0.240)	(0.237)	(0.173)
School size	0.531***	0.384***	0.483***
	(0.030)	(0.047)	(0.031)
School size squared	-0.031***	-0.029***	-0.026***
	(0.002)	(0.003)	(0.002)
Ppal. Enhance reputa	-0.269***	-0.432***	-0.135**
	(0.055)	(0.079)	(0.057)
Prop Dropout	0.054***	-0.004	0.071***
	(0.005)	(0.007)	(0.005)
Dropout75	-1.037***	-0.066	-1.282***
	(0.115)	(0.172)	(0.122)
Constant	-1.489***	-1.496***	-1.704***
	(0.133)	(0.174)	(0.109)
Observations	11,025	8,811	10,229

Note: The probability of participating in the program is estimated using a Logit model which, in addition to the covariates shown in the table, includes the following control variables: regional dummies (Andalusia, Aragon, Castile Leon, Catalonia, Extremadura, Galicia, Murcia and Navarre) and dummies to capture missing values in some variables (attended pre-primary, mother educated and school size). *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis. Source: PISA 2012

Table 8: The impact of PAE

	PAE-Immediacy	PAE-Intensity	
		1-2 Years	3-4 Years
		Panel A: Reading25	
IPWEnc	-0.033*	-0.026	-0.032***
	(0.019)	(0.023)	(0.020)
IPWEwc	-0.030**	-0.022	-0.031**
	(0.013)	(0.018)	(0.016)
NNPS(2)	-0.059***	-0.039**	-0.075***
	(0.012)	(0.016)	(0.015)
NNPS(4)	-0.041***	-0.026*	-0.063***
	(0.011)	(0.014)	(0.013)
NNPS(6)	-0.044***	-0.028**	-0.058***
	(0.010)	(0.014)	(0.013)
NNPS(8)	-0.044***	-0.028**	-0.051***
	(0.010)	(0.013)	(0.012)
		Panel B: Reading	
IPWEnc	6.156	1.020	6.127
	(4.523)	(5.415)	(4.845)
IPWEwc	5.530*	1.030	5.962
	(3.200)	(4.294)	(3.745)
NNPS(2)	12.343***	6.14*	16.18***
	(2.344)	(3.18)	(2.99)
NNPS(4)	8.899***	2.579	12.70***
	(2.110)	(2.93)	(2.671)
NNPS(6)	8.930***	2.720	11.867***
	(2.03)	(2.784)	(2.506)
NNPS(8)	8.645***	2.180	10.747***
	(1.990)	(2.710)	(2.400)
Observations	11,025	8,811	10,229

Note: The first two rows in Panel A and Panel B report Inverse Probability Weighting Estimator without covariates (IPWEnc) or with covariates (IPWEwc). The covariates included in the estimation are the ones used for the propensity score (Table 7); Rows 3 to 6 in Panel A and Panel B report nearest neighbour propensity score estimators using two neighbours NNPS(2), four NNPS(4), etc. *** p<0.01, ** p<0.05, * p<0.1. Robust standard standard errors in parenthesis. Standard errors for the IPWE are corrected for clustering at the school level. Source: PISA 2012.

Table 9: Estimated CDF reading scores

Percentile	Value	Estimated CDF			Increase Treated Weighted Control
		Control	Weighted Control	Treated	
Panel A: PAE-Immediacy					
5	321	0.042	0.054	0.054	-0.002
10	359	0.083	0.107	0.099	-0.077
15	384	0.125	0.161	0.145	-0.097
20	404	0.171	0.211	0.195	-0.079
25	420	0.214	0.263	0.234	-0.110
30	434	0.258	0.308	0.277	-0.099
40	458	0.341	0.398	0.372	-0.065
50	482	0.443	0.500	0.480	-0.040
60	505	0.551	0.609	0.579	-0.049
70	527	0.652	0.702	0.680	-0.032
80	553	0.761	0.803	0.789	-0.018
90	586	0.880	0.904	0.890	-0.016
Panel B: PAE-Intensity: 1-2 years					
5	321	0.042	0.049	0.0519	0.052
10	359	0.083	0.098	0.103	0.056
15	384	0.125	0.148	0.151	0.017
20	404	0.171	0.204	0.194	-0.049
25	420	0.214	0.254	0.233	-0.083
30	434	0.258	0.306	0.288	-0.060
40	458	0.341	0.399	0.389	-0.024
50	482	0.443	0.505	0.502	-0.006
60	505	0.551	0.613	0.613	-0.000
70	527	0.652	0.708	0.713	0.008
80	553	0.761	0.809	0.814	0.006
90	586	0.880	0.907	0.903	-0.005
Panel C: PAE-Intensity: 3-4 years					
5	321	0.042	0.053	0.052	-0.013
10	359	0.083	0.107	0.098	-0.080
15	384	0.125	0.162	0.146	-0.098
20	404	0.171	0.215	0.198	-0.080
25	420	0.214	0.268	0.240	-0.104
30	434	0.258	0.317	0.284	-0.105
40	458	0.341	0.408	0.382	-0.065
50	482	0.443	0.506	0.488	-0.035
60	505	0.551	0.614	0.582	-0.052
70	527	0.652	0.708	0.681	-0.037
80	553	0.761	0.803	0.795	-0.011
90	586	0.880	0.904	0.892	-0.013

Note: Columns 3, 4 and 5 show the CDF of reading scores among control students, re-weighted controls students and among treated students, respectively. Column 6 presents the difference between columns 4 and 5 (rate equal to the CDF treated/CDF weighted controls minus one).

Table 10: The impact of the PAE program: subsamples

	PAE-Immediacy		PAE-Intensity			
			1-2 years		3-4 years	
	P<50	P>50	P<50	P>50	P<50	P>50
	Reading25	**Reading75**	**Reading25**	**Reading75**	**Reading25**	**Reading75**
IPWEnc	-0.045**	0.017	-0.039	-0.015	-0.042*	0.018
	(0.022)	(0.020)	(0.027)	(0.029)	(0.024)	(0.022)
IPWEwc	-0.041**	0.018	-0.032	-0.013	-0.046**	0.017
	(0.019)	(0.018)	(0.022)	(0.027)	(0.022)	(0.020)
NNPS(2)	-0.051***	0.030	-0.046*	0.006	-0.024	0.046**
	(0.019)	(0.019)	(0.026)	(0.029)	(0.022)	(0.020)
			Reading			
IPWEnc	3.388	3.309*	0.725	-0.042	3.180	3.634
	(3.103)	(1.944)	(3.821)	(2.827)	(3.255)	(2.210)
IPWEwc	2.614	3.347*	0.150	0.049	3.438	3.464*
	(2.559)	(1.827)	(3.055)	(2.488)	(2.987)	(2.023)
NNPS(2)	3.045	5.106***	2.273	0.046	2.337	5.530***
	(2.339)	(1.727)	(3.143)	(2.543)	(2.725)	(1.816)
Observations	4,990	6,035	3,964	4,847	4,632	5,597

Note: P<50 (resp. P>50) refers to the subsample of students below (above) the median of the reading distribution for the sample of all public schools. Reading75 indicates the probability of having a reading score above the third quartile of the reading distribution. *** p<0.01, ** p<0.05, * p<0.1. Robust standard standard errors in parenthesis. Standard errors for the IPWE are corrected for clustering at the school level. Source: PISA 2012.

Table 11: The impact of PAE program: Urban vs rural

	PAE-Immediacy		PAE-Intensity			
			1-2 YEARS		3-4 YEARS	
	Urban	**Rural**	**Urban**	**Rural**	**Urban**	**Rural**
	Panel A: Reading25					
IPWEnc	-0.0150	-0.0631**	-0.0439	-0.0275	-0.0153	-0.0584**
	(0.0246)	(0.0255)	(0.0301)	(0.0337)	(0.0277)	(0.0293)
IPWEwc	-0.01072	-0.0613***	-0.0262	-0.0210	-0.0187	-0.0578***
	(0.01102)	(0.0200)	(0.0234)	(0.025)	(0.0217)	(0.022)
NNPS(2)	-0.03555**	-0.0746***	-0.0849**	-0.0776***	-0.0378**	-0.0979***
	(0.01599)	(0.0194)	(0.0397)	(0.0238)	(0.0184)	(0.0302)
	Panel B: Reading					
IPWEnc	2.1982	14.7637**	1.593	3.491	2.8154	14.2214*
	(6.0298)	(6.4154)	(7.557)	(7.889)	(6.4793)	(7.5544)
IPWEwc	2.1341	13.7680***	0.1228	3.025	3.8413	14.4464***
	(2.0918)	(4.9578)	(3.0221)	(5.542)	(5.0384)	(5.547)
NNPS(2)	10.0955***	16.2528***	13.6478**	14.7057***	6.7055**	24.9540***
	(3.1401)	(3.8122)	(5.4683)	(4.3500)	(3.0991)	(5.8178)
Observations	6,272	4,549	4,818	3,857	5,879	4,171

Note: Urban (resp. rural) schools are those located in a community of more (resp. less) than 15,000 people. *** p<0.01, ** p<0.05, * p<0.1. Robust standard standard errors in parenthesis. Standard errors for the IPWE are corrected for clustering at the school level. Source: PISA 2012.

Table 12: Summary Statistics: Treated and controls. Selection bias

	Treated	Control	Weighted control
PISA scores			
Reading25	0.274	0.217	0.271
Reading	472.7	488.0	472.8
Individual variables			
Gender (girl)	0.481	0.497	0.470
Immigrant	0.151	0.069	0.164
Repeater once	0.303	0.222	0.305
Repeater more	0.095	0.07	0.104
Attended pre-primary	0.797	0.825	0.803
Socioeco background			
Father educated	0.588	0.635	0.578
Mother educated	0.593	0.679	0.591
Index educ possesions	-0.260	-0.104	-0.297
School variables			
Students educ parents	0.417	0.527	0.51
ESCS	-0.455	-0.197	-0.256
Presion	0.487	0.407	0.398
School size	549.5	618.4	577.4
Prop Immigrants	0.13	0.113	0.116
Student Teacher Ratio	7.236	8.634	7.925
Rural	0.424	0.386	0.418
Observations	912	3,656	3,488

Note: Treated: students attending schools that did not participate in the PAE program before 2009 but participated after 2009 (either during 2009/10, 2010/11 or 2011/12). Treated and control in columns (1) and (2) are sample averages. Weighted controls are averages of the control group when the sample is reweighted by the (inverse of the) probability of participating in PAE only after 2009 predicted by the set of individual, socioeconomic and school variables in this table. Source: PISA 2009

Table 13: Determinants of PAE participation only after 2009

	PAE only after 2009
Individual variables	
Gender (girl)	-0.024
	(0.087)
Immigrant	0.621***
	(0.143)
Immigra x Murcia	0.468
	(0.420)
Repeater once	0.380***
	(0.103)
Repeater more once	0.428***
	(0.162)
Attended pre-primary	0.306***
	(0.118)
Socioeconomics variables	
Mother educated	-0.398***
	(0.119)
Mother educ x BasqueCountry	0.746***
	(0.192)
Father educated	0.021
	(0.103)
Index educ pos	-0.087*
	(0.049)
School variables	
ESCS	-0.386***
	(0.119)
School size	-0.042***
	(0.016)
Students educ parents x Presion	0.003*
	(0.002)
Stu Teach Ra x Ast	0.135***
	(0.035)
Stu Teach Ra x Canta	-0.132**
	(0.057)
Stu Teach Ra x Basque Country	-0.104***
	(0.038)
Rural	0.367***
	(0.108)
Constant	-4.339***
	(0.332)
Pseudo R-squared	0.166
Observations	4,314

Note: The probability of participating in PAE program only after 2009 is estimated using a logit model model which, in addition to the covariates shown in the table, includes the following control variables: regional dummies (Aragon, Balearic Islands, Cantabria, Galicia, La Rioja, Murcia and Basque Country) and dummies to capture missing values in some variables (attended pre-primary, mother and father educated). *** p<0.01, ** p<0.05, * p<0.1, Robust standard errors in parenthesis. Source: PISA 2009

Table 14: Impact of PAE participation (only after 2009)

Reading25	
OLS	0.009
	(0.035)
IPWEnc	0.002
	(0.042)
IPWEwc	0.014
	(0.039)
NNPS(2)	-0.002
	(0.029)
Reading	
OLS	1.077
	(7.966)
IPWEnc	0.825
	(9.374)
IPWEwc	-0.178
	(8.134)
NNPS(2)	2.915
	(4.26)
Observations	4,314

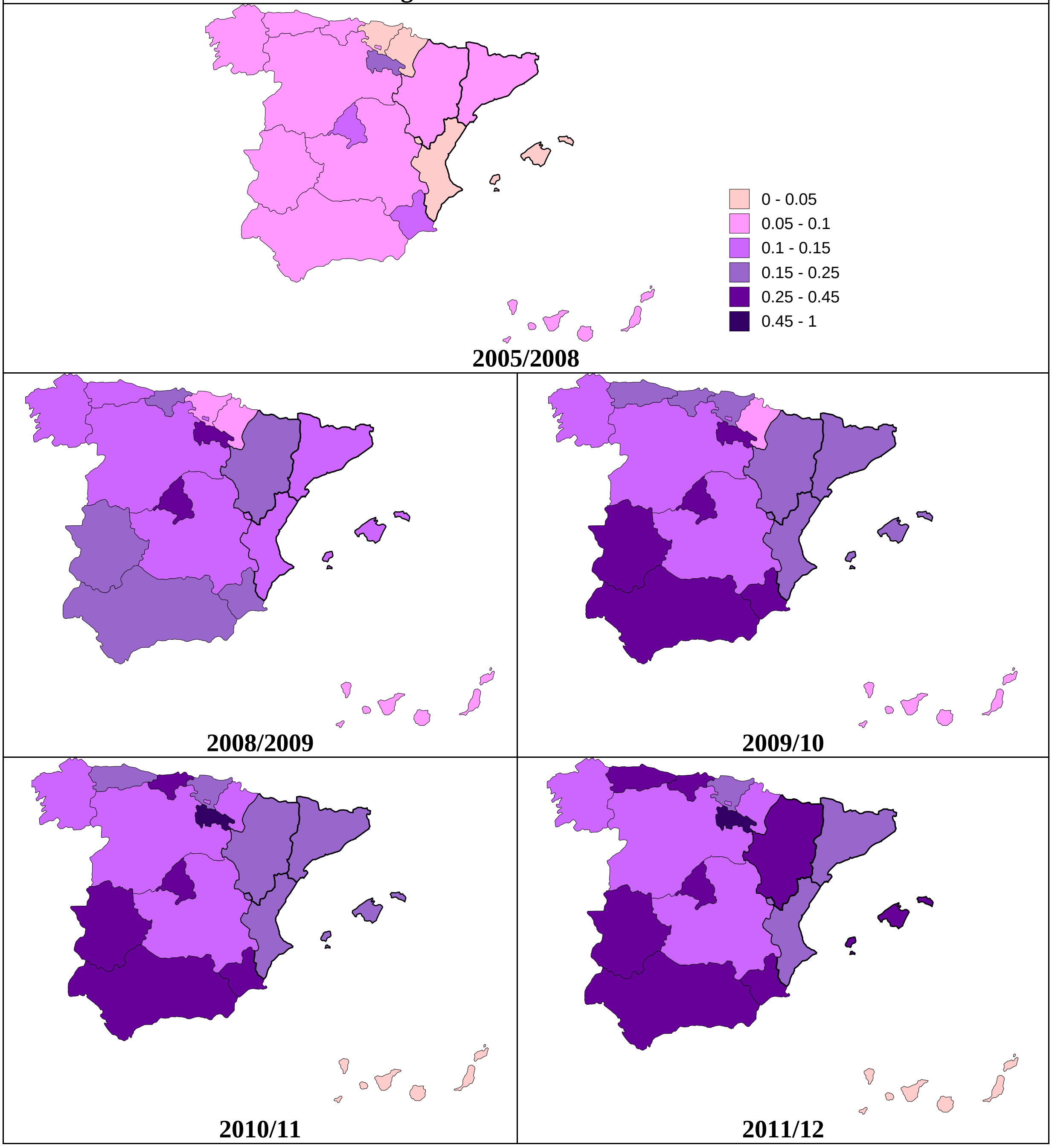
Note: The dependent variable is the student's probability of belonging to the first quartile of the reading distribution in the PISA 2009 exams for public schools (Reading25) and the student's reading score in the PISA 2009 exams. The estimation method in the first row is ordinary least squares. Estimation methods in the rest of rows are similar to the ones used in Tables 8 and 10 above. *** p<0.01, ** p<0.05, * p<0.1, Robust standard errors are clustered at the school level.

Table 15: The impact of the PAE program: schools

	PAE-Immediacy	PAE-Intensity	
		1-2 Years	3-4 Years
		Panel A: Reading25	
IPWEnc	-0.0615*	-0.0316	-0.0574
	(0.0362)	(0.0427)	(0.0378)
IPWEwc	-0.0485***	-0.0121	-0.0506***
	(0.0157)	(0.0196)	(0.0181)
NNPS (2)	-0.0757**	-0.0493	-0.0809***
	(0.0310)	(0.0305)	(0.0242)
NNPS (4)	-0.0533**	-0.0370	-0.0663**
	(0.0269)	(0.0242)	(0.0277)
NNPS (6)	-0.0490**	-0.0093	-0.0533
	(0.0206)	(0.0080)	(0.0209)
NNPS (8)	-0.0423***	-0.0154	-0.0431***
	(0.0149)	(0.0047)	(0.0163)
		Panel B: Reading	
IPWEnc	11.7914	6.9013	11.7294
	(8.3026)	(10.5867)	(9.0359)
IPWEwc	9.2289**	2.2017	10.3672**
	(4.2101)	(4.5658)	(4.7160)
NNPS (2)	12.1297**	6.4185	19.7893***
	(5.6719)	(6.4560)	(5.8397)
NNPS (4)	10.3012	5.1868	13.1461***
	(7.4340)	(5.9163)	(3.7609)
NNPS (6)	9.7147	-1.1163	10.5400***
	(6.1936)	(3.8850)	(3.4331)
NNPS (8)	8.3082*	-0.3983	8.0727***
	(5.0241)	(2.9315)	(2.4933)
Observations	395	317	366

Note: The first two rows in Panel A and Panel B report Inverse Probability Weighting Estimator without covariates (IPWEnc) or with covariates (IPWEwc). The covariates included in the estimation are the ones used for the propensity score (Table 7); Rows 3 to 6 in Panel A and Panel B report nearest neighbour propensity score estimators using two neighbours NNPS(2), four NNPS(4), etc. *** p<0.01, ** p<0.05, * p<0.1. Robust standard standard errors in parenthesis. Standard errors for the IPWE are corrected for clustering at the school level. Source: PISA 2012

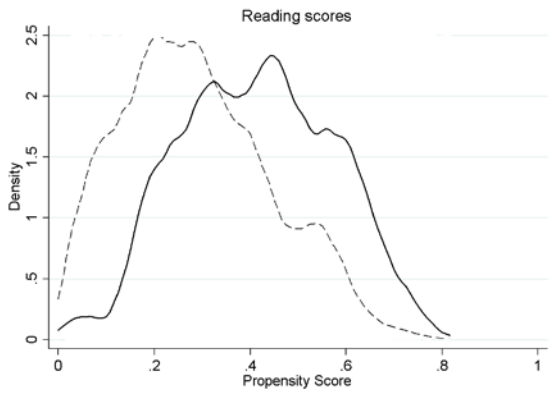
Figure 1: Schools with PAE



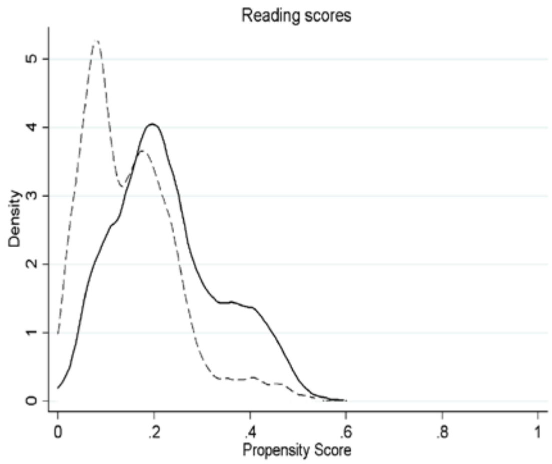
Note: Percentage of PAE secondary schools over all public secondary schools. Source: INEE and Spanish Ministry of Education (2016)

Figure 2: Propensity score support

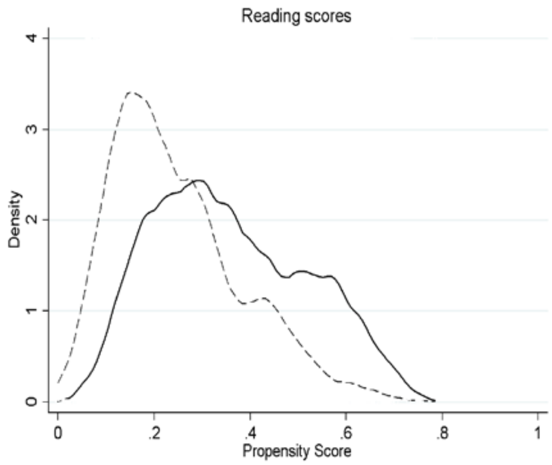
PAE-Immediacy



PAE-Intensity: 1-2 years



PAE-Intensity: 3-4 years



Note: Density of the probability of participation (propensity score) for treated (solid line) and control groups (dotted line). The figure in the upper part corresponds to the propensity score computed for the PAE-Immediacy treatment (Table 7 column 1). The figures in the center and bottom part correspond to the propensity score computed for the PAE-Intensity, 1-2 years and 3-4 years, respectively (Table 7 columns 2 and 3).

Appendix 2

Table A.1: Summary Statistics: treated and control. Urban vs rural schools

	Urban schools			Rural schools		
	Treated	Control	Weighted Control	Treated	Control	Weighted Control
Number of schools	74	146	144	55	120	116
Number of students	2,181	4,275	4,175	1,485	3,184	3,106
PISA scores						
Reading25	0.230	0.179	0.231	0.241	0.264	0.297
Reading	481.9	498.1	483.2	476.9	472.5	463.9
Individual variables						
Gender (girl)	0.494	0.506	0.492	0.507	0.510	0.507
Immigrant	0.190	0.0957	0.185	0.102	0.0782	0.0991
Repeater once	0.271	0.208	0.265	0.271	0.247	0.276
Repeater more once	0.126	0.0802	0.121	0.136	0.101	0.136
Attended pre-primary	0.802	0.829	0.815	0.828	0.829	0.836
Socioeconomic Variables						
Father educated	0.311	0.410	0.304	0.275	0.262	0.271
Mother educated	0.315	0.418	0.310	0.281	0.295	0.279
Index of educ pos	0.0235	0.0964	0.0449	0.0662	0.0434	0.0602
School variables						
Stu Teacher Ratio	8.846	10.22	8.827	8.101	7.683	7.892
ESCS	0.332	0.612	0.346	0.213	0.221	0.190
School size	6.912	6.878	6.957	4.391	3.831	4.356
Prop Dropouts	11.64	8.986	10.08	11.54	7.940	12.02

Note: Treated students under Treatment PAE-Immediacy. Treated and control in columns (1) and (2) and (4) and (5) are sample averages. Weighted controls in columns (3) and (6) are averages of the control group when the sample is reweighted by the (inverse of the) probability of receiving PAE-Immediacy treatment predicted by the set of individual, socioeconomic and school variables in this table. Source: PISA 2012

Table A.2: Summary Statistics: Evaluation sample and PISA sample. Schools

	Evaluation sample	All public schools
Reading scores		
Mean	475.32	476.36
Standard Deviation	82.12	83.39
School Variables		
Prop. repeater once	0.207	0.243
Prop. repeater more	0.115	0.101
Students educ parents	0.174	0.168
ESCS	-0.348	-0.390
School size	575.96	580.49
Prop Immigrants	0.106	0.114
Prop Dropout	0.102	0.108
Student Teacher Ratio	10.15	9.94
Rural	0.408	0.378
Ppal Admittance	0.385	0.404
Ppal Staff Decision	0.759	0.763
Ppal Enhance Reputat.	0.247	0.255
Ppal Review Work	0.192	0.191
Observations	417	543

Note: Evaluation sample: PISA sample excluding students in private schools and schools which joined other remedial programs. See the text for details. Source: PISA 2012

Table A.3: Summary Statistics: Treated and control. Schools

	Treated	**Control**	**Weighted control**
PISA scores			
Reading	471.8	476.4	458.1
Reading25	0.263	0.271	0.350
School Variables			
Prop. repeater once	0.310	0.225	0.316
Prop. repeater more	0.155	0.113	0.122
Students educ parents	0.163	0.210	0.161
ESCS	0.263	0.406	0.259
School size	568.1	531.7	562.8
Prop Immigrants	0.185	0.102	0.172
Prop Dropout	0.127	0.092	0.125
Student Teacher Ratio	8.43	8.93	8.578
Rural	0.426	0.451	0.463
Ppal Admittance	0.364	0.327	0.367
Ppal Staff Decision	0.636	0.793	0.693
Ppal Enhance Reputat.	0.217	0.237	0.230
Ppal Review Work	0.163	0.143	0.160

Note: Treated schools under Treatment PAE- Immediacy. Treated and control in columns (1) and (2) are sample averages. Weighted controls in columns (3) are averages of the control group when the sample is reweighted by the (inverse of the) probability of receiving PAE-Immediacy treatment predicted by the set of school variables in this table. Source: PISA 2012

Table A.4: Propensity score estimation. Schools

	PAE-Immediacy	PAE-Intensity	
		1-2 years	3-4 Years
Repeaters	2.1184**	1.8333*	2.3184**
	(0.8421)	(1.0588)	(0.9210)
Students educ parents	-0.0098	-0.0042	-0.0124
	(0.0140)	(0.0209)	(0.0159)
Prop Immigrants	0.2235	-0.2941	0.5048
	(0.3489)	(0.4976)	(0.3806)
ESCS	-0.4116	-1.0479*	-0.2793
	(0.3799)	(0.5695)	(0.4356)
School size	0.7676***	0.6160**	0.7784***
	(0.1814)	(0.2834)	(0.1995)
Prop Dropout	0.0023	-0.0050	0.0070
	(0.0117)	(0.0164)	(0.0129)
Stu Teach Ratio	-0.1925***	-0.16836*	-0.1553*
	(0.0738)	(0.1019)	(0.0821)
Rural	0.1808	-0.2083	0.2219
	(0.3123)	(0.4002)	(0.3431)
Ppal Stu Admittance	0.4006	0.5980	0.4965
	(0.2879)	(0.3795)	(0.3103)
Ppal Staff Decision	-0.9612***	-1.0647***	-0.9931***
	(0.2866)	(0.3932)	(0.3159)
Ppal Enhance Repu	-0.3064	-0.4114	-0.1835
	(0.3103)	(0.4276)	(0.3318)
Ppal Rev Stu Work	0.2681	0.4523	0.4481
	(0.3384)	(0.4622)	(0.3641)

Note: The probability of participating in the program is estimated using a Logit model which, in addition to the covariates shown in the table, includes the following control variables: regional dummies (Andalusia, Aragon, Asturias, Balearic Islands, Cantabria, Castile Leon, Catalonia, Extremadura, Galicia, La Rioja and Navarre) and dummies to capture missing values in some variables (school size). *** p<0.01, ** p<0.05, * p<0.1, Standard errors in parenthesis. Source: PISA 2012